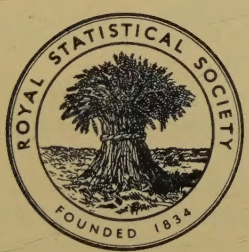
1486 60

APPLIED STATISTICS

A Journal of the
Royal Statistical Society

EDITED BY
LEONARD H. C. TIPPETT



VOLUME ONE

Oliver and Boyd Ltd.
London: 98 Great Russell Street, W.C.1
Edinburgh: Tweeddale Court

1952

CONTENTS

Page

Foreword — *A. Bradford Hill, C.B.E., President, Royal Statistical Society* 1

ARTICLES

The Introduction of Statistical Methods to Industry — *Bernard P. Dudding* 3
A Statistician looks at the International Management Congress of 1951 — *Albert W. Swan* 21
Estimating the Social Class of Towns .. .. — *Leslie T. Wilkins* 27
A Use for the Jury Qualification in Sample Design — *Thomas Corlett* 34
A Statistical Approach to the Specification of Plastics — *Charles Wainwright* 37
Uses and Abuses of Factor Analysis .. .. — *Hans J. Eysenck* 45
The Accuracy of Systematic Sampling from Conveyor Belts — *Geoffrey H. Jowett* 50
Estimating the Future Population .. .. .. — *Peter R. Cox* 82
A Collaborative Investigation into Methods of Chemical Analysis of Ground Pyrethrum Flowers — *George V. Dyke and Emily P. Simpson (née Poulton)* 95
Index Numbers of Retail Prices, 1938–51 .. — *Roy G. D. Allen* 106
Calibration of an Integrating Gas Flowmeter .. — *William Spendley* 118
The Measurement of Branch Sales Performance — *Bernard T. Ramm* 126
One Line or Two? .. .. — *William N. Jessop* 131
The Scope of Medical Statistics .. .. .. — *Norman T. J. Bailey* 149
A Study of the Lives of Pots used in making Glass — *Joan Keen and Denys J. Page* 163
The Experience of a User of a Consumer Panel — *Robert N. Wadsworth* 169
The Accuracy of Sampling Coal — *Jack Hebden and Geoffrey H. Jowett* 179
Estimating Individual Expenditure from Family Totals — *W. F. F. Kemsley* 192
Composite Mean Squares and their Degrees of Freedom — *LeRoy P. V. Johnson and Ernest S. Keeping* 202

MISCELLANEA

Editorials 2, 81
Questions and Answers 60, 206
Notes and Comments 63, 138, 208
Industrial Applications Section of the Royal Statistical Society: Reports 65, 140, 210
Study Section of the Royal Statistical Society: Reports .. 74, 145, 211
Book Reviews 80, 148
Index

Erratum

p. 87, *Table I, last column. For* '− 0·17' *read* '+ 1·83'.

FOREWORD

THE world today is full of scientific journals, so full that many a scientific worker requires a second set—of abstracting journals—to enable him to keep pace with the 'literature' of his subject. Anyone bold enough to found a new one must therefore have a clear and unanswerable case. In founding *Applied Statistics* the Council of the Royal Statistical Society felt sure not only that they had such a case but that, in fact, they had a duty to meet.

The Council knew that throughout the post-war years there had been a growing demand, within the Society and without, for a publication devoted to the practical statistical problems that arise in so many fields of human activity. For every person interested in theoretical statistics, and so fortunately endowed as to be able to follow the mathematical-statistical trends of the present time, there are many whose interests lie in the application of these newly developed methods. For many a statistician the main problems are those involved in the collection of accurate data and—most difficult of all—in their presentation clearly, simply, and effectively. With these aspects of statistics indeed every practising statistician is concerned, whether he be mathematically minded or not.

Applied Statistics has been founded, therefore, to meet the needs of all workers concerned with statistics—not of professional statisticians only but also of those innumerable workers in industry, commerce, science, and other branches of daily work, who must handle and understand statistics as part of their tasks. Its aim, in short, is to present, in one way or another but always simply and clearly, the statistical approach and its value, and to illustrate in original articles modern statistical methods in their everyday applications.

In its endeavour to fulfil this aim the Council is most fortunate in having secured Mr. L. H. C. Tippett of the British Cotton Industry Research Association as the first editor. It is my privilege, on behalf of the Council, to wish him success in his task and the journal a long and useful life. I shall not commend it to a wide circle of readers; *Applied Statistics* must gain that circulation by its own appeal and upon its own merits. I have no doubt myself that it will do so.

A. BRADFORD HILL

President

Royal Statistical Society

EDITORIAL

THIS first number of *Applied Statistics* shows the main features of the journal, but clearly no single number can show them all. There will continue in future numbers to be articles of general interest for the layman as well as technical articles, which will usually be shorter; but the subjects and fields of application will be widened. This first number has a bias towards commerce and industry, and those interests are always likely to be strongly represented; but we hope also to publish articles to interest economists and social scientists, medical scientists and agricultural scientists, chemists and physicists, engineers and technologists. One of the fascinations of statistics is that its applications bring one into contact with so wide a range of subjects and interests.

It is, of course, in the hands of readers to make 'Questions and Answers' and 'Letters to the Editor' lively features. For this first number questions have been sent by members of the editorial committee who have also supplied the answers, and there has been no opportunity for letters. We shall welcome, not only questions and letters for publication in future numbers, but also letters of suggestion and criticism not intended for publication.

That the substance of articles accepted for publication in *Applied Statistics* must have quality goes without saying, but the form in which they appear also matters. The publishers and printers have produced a journal that, in our opinion, is attractive to the eye; and we on the editorial side aim at maintaining a sound literary quality. Not everyone finds it easy to construct a well-balanced article, with everything in the best order and with just the right amount of detail and explanation; nor has everyone the gift of being able to write good, clear English easily. In particular, the technical man writing on a technical subject often finds it difficult to avoid using too much technical jargon. We shall take pains in this matter and shall ask contributors to take pains, even though that may sometimes involve revising or redrafting articles. To produce something that is easy to read and understand is worth taking a lot of trouble.

This number contains 80 pages. For the time being we shall aim normally at 64 pages, hoping that ultimately the larger size will become regularly possible.

THE INTRODUCTION OF STATISTICAL METHODS TO INDUSTRY

BERNARD P. DUDDING

Vice-chairman, Advisory Scientific Panel, Research Laboratories
The General Electric Co. Ltd., England

Dr. Dudding is a pioneer in this country in the application of statistics to the investigation and control of factory processes. Here he describes how he and his colleagues at the GEC developed this application to its culmination in the use of the control chart technique. The article also deals with practical points that arise in introducing this technique into a factory and describes the advantages that accrue from its application.

Introduction

Reference to the very extensive bibliography provided by Professor M. G. Kendall in his recent book *The Advanced Theory of Statistics*[1] shows that considerable attention was being given to the development of statistical theory in the early years of the present century. However, the large majority of engineers, physicists, and chemists remained unaware of this development. There were two reasons for this: (i) the papers were published in journals which would not be read by them, particularly if they were engaged in industry, and (ii) even the most elementary statistical ideas found no place in their education at technical colleges and universities, a state of affairs which is being remedied far too slowly in Great Britain even today.

When I first became interested in statistical methods in 1913 my knowledge of the subject was derived from an early edition of 'Udny Yule' and my attention was never directed to the theoretical papers referred to above.

The systematic application of statistical methods by persons engaged in research work appears to have begun about 1920. The available literature shows that the most important contributions during the 1920's were stimulated by the work of Professor R. A. Fisher and that statistical methods had begun to play an effective part in agricultural research and thereby to have a beneficial effect on the agricultural industry of Great Britain.

In 1924 publications in the *Bell System Technical Journal*[2,3] showed that statistical methods were being used in the telephonic communications industry of the United States of America.

At this time my colleagues and I in the Research Laboratories of the General Electric Co. Ltd., which was founded in 1916 and became operative in 1919, had begun to use statistical methods in our studies of the manufacturing processes of lamps and of the metals and glasses

used in their construction. In 1919 the inspection of the final products of many processes followed traditional methods: methods which still govern the inspection of a high proportion of manufactured products in most industries. The inspection frequently involved the cost of the final examination of every item made and the cost of rectifying or scrapping unsatisfactory material and articles. Of course, where the tests were destructive, samples were selected from bulked products, but the number of items in a sample was determined arbitrarily or followed established custom and was not determined, as it should have been, by consideration of the variability present in the bulk sampled and of the accuracy of the result which technical considerations demanded.

As scientific and technical knowledge of the processes became available the older systems of inspection were gradually replaced by routine inspection of samples at those stages of production which this knowledge and manufacturing experience had demonstrated were of prime importance. The tests on samples were associated with the graphical presentation of the results on charts whose construction was based on statistical principles and methods. These methods made important contributions to a continuous reduction in the waste of material and labour and to a steady improvement in the quality of the commercial product.

Encouraged by the late Sir Clifford Paterson and by the Directors of the GEC, some of my colleagues and I have published from time to time since 1926 papers illustrating the techniques (see bibliography), and during the last fifteen years we have given lectures in universities and technical colleges to students, and in factories to many hundreds of persons engaged in a wide diversity of industries. In discussions following these lectures I have frequently been asked how I became interested in the use of statistical methods. While it is difficult to recall the steps in the evolution of a technique of this kind I will try to do so, and at the same time indicate how the technique assists those engaged in industry to raise production efficiency. By so doing I hope to encourage a higher proportion of those engaged in factory supervisory duties to acquire familiarity with statistical modes of thought and techniques, and I aim to stimulate senior executive officers to encourage members of their staffs to use fully the steadily increasing facilities for doing so.

These facilities are of two kinds: (i) courses of lectures and instruction at educational establishments, and (ii) meetings arranged for reading and discussing papers on the methods of applied statistics where production engineers and technicians of all grades can exchange information concerning their experiences with the practical problems which arise. Whatever the industry, these problems have much in common. The meetings also provide opportunities for exchange of views with trained statisticians and for engaging the interest of the latter in the application of statistical methods to industrial problems.

HISTORICAL

I became interested in statistical methods soon after accepting in 1911 a post at the National Physical Laboratory in the photometric department as assistant to the late Sir Clifford Paterson. Among the duties of a small group was an investigation directed towards the establishment of an international unit for the measurement of light. It was the analysis of the data accumulated by six observers, who made observations nearly every day for two years on the light output of the standard pentane lamp and on one of six groups of incandescent tungsten filament lamps, that aroused my interest in statistical methods.

The light output of the pentane lamp is influenced by the height of the flame, and the pressure, humidity and temperature of the atmosphere. The colour of the light approximates to that of an incandescent tungsten filament operating at about 1600° C. The filaments of the lamps in common use at that time operated at about 2000° C. and the light appeared much bluer than the light of the pentane lamp when making visual photometric comparisons by the contrast method—the most sensitive method then available. This difference in colour introduced differences in the results of observations made by any one observer on the same day, larger differences between the results for the same observer on different occasions, and even larger differences between observers. The investigation was therefore divided into two main sections:

(*a*) The comparison of the light output of the standard pentane lamp with the average of about twelve tungsten filament lamps designed to give light of the same colour as the pentane lamp.

(*b*) A comparison of these foregoing tungsten filament lamps with a similar group of tungsten filament lamps designed to give light whose colour approximated to that of the lamps commercially available. This comparison was carried out in two ways: (i) by direct comparison, the maximum colour difference being present in all comparisons, and (ii) by using four intermediate groups of lamps designed to divide the colour difference into approximately equal steps. The investigation is fully described in a paper published in 1915.[4]

The principal statistical interest in this paper does not reside in the regression which connects the candle-power of a pentane lamp with the pressure, the humidity, and the temperature of the surrounding air, but rather in the attempt made to estimate the magnitude of the contribution from different sources to the error in the final result. These contributions arise from the differences between the observations of an observer on the same day, differences between the observations of the same observer on different days, and differences between the observations of different observers. In the last two groups are included the differences which arose when adjusting the height of the flame of the pentane lamp.

During the first world war I was in charge of a small department

devoted to the testing of screw gauges. While this work was important in itself, the more valuable work of this period was done at week-ends in workshops devoted to gaugemaking. The many hours I spent talking to skilled mechanics, chargehands, and foremen, explaining the principles of new measuring devices and promoting their adoption in the workshop, not only helped to solve what appeared in 1915 to be an intractable problem, but also marked a further stage in my own education.

Although I agreed early in 1916 to join Clifford Paterson when he undertook the task of founding the Research Laboratories of the General Electric Company, no practical steps were taken during the next two and a half years. Arising out of my industrial experiences during that time we occasionally discussed the problems other than scientific and technical which would arise when we were actually engaged in providing a scientific service to an industry which was considered at that time to be both technically progressive and commercially successful. January 1st, 1919, found my late chief and me, together with a secretary, housed in temporary accommodation at the Osram Lamp Works at Hammersmith, and a great adventure was on.

The first task allotted to me was the reorganisation and control of the testing arrangements for lamps, which had been a feature of the industry from the earliest days of the carbon filament lamp, introduced in 1880. I doubt whether this somewhat routine task would have appealed to the average graduate in engineering or physics of that period any more than the introduction of the control chart technique in a factory would appeal to the fully trained graduate in statistics today; but the position proved to be a key one throughout the following ten to fifteen years.

In 1919 every lamp was measured for power consumption and light output before it was marked with the appropriate voltage and wattage rating. In addition, quite large samples of lamps, selected at random weekly from the manufactured product, were placed on life test in order to obtain information on the performance of the lamps in respect to maintenance of light output (or efficiency) and length of life. The measurement of all the lamps involved costly apparatus which occupied a large amount of floor space, and loss of product due to handling. Later research investigations showed that the heating of the glass envelope which was necessary to satisfactory marking caused loss of lamp quality.

The life testing also involved the use of expensive apparatus (see plate), considerable expenditure on power (the cost of which amounted to about ten times the value of the lamp), and, of course, the destruction of several thousands of lamps per annum. It is also important to note that if the results were unsatisfactory it was at first rarely possible to identify by inspection which of the untested lamps were likely to prove unsatisfactory. However, in a year or two as scientific and technical knowledge grew it became possible to eliminate some of the worst

A photograph of equipment for testing electric lamps.

lamps by inspecting batches of lamps after operating them for periods up to 50 hours. This was very inconvenient and costly and not always effective. On occasions the only method of eliminating poor lamps was to scrap large batches of lamps—a drastic procedure followed only in extreme cases.

Introduction of Statistical Methods to Factory Personnel

During the years 1919, 1920, and 1921, the scientific staff of the Laboratories investigating lampmaking processes was increased from two to eight. We were engaged on investigations designed to reduce loss of material at the various stages of lamp assembly. These consisted broadly of two kinds: investigations of the physics and chemistry of some of the lamp assembly processes and investigations designed to improve the quality of the glass and metals used in the manufacture of lamp components. Any improvement suggested by our work had ultimately to be demonstrated by either an improvement in the quality of the lamps or a reduction in material lost during their manufacture. It was during these years, therefore, that the problem arose of comparing the quality of a few lamps made under modified conditions or with materials prepared by new methods, with the quality of lamps made by following the usual factory routine. In order to make a comparison of this kind it was necessary to devise a method by which the quality of the factory product could be expressed.

The routine measurements made on samples selected from the lamps manufactured each week were: (i) a measurement of the initial efficiency of the lamp when operating at its marked voltage, (ii) the length of life, and (iii) the average efficiency throughout life when operated continuously at the marked voltage. The practice was established in 1920 of summarising the results of all these tests in an annual report. These reports at first followed a traditional method of presentation, the average results being given in tabular form and the only reference to spread in quality being the proportion of the lamps tested which failed before 500 hours—the designed average life being in the neighbourhood of 1200 hours.*

As an initial step towards defining statistically the quality of the production, the results of the routine tests were prepared as a composite histogram, the results for the lamps made each month being represented by a different symbol. A diagram of this type is illustrated in Fig. 1. The results of the initial efficiency tests and of the length of life tests were presented in this way and the method found favour with the supervisory staffs in the factory. At the same time the length of life data were presented as a 'survivor' curve because this enabled the proportion

* The first British Standard Specification for tungsten filament lamps was issued in June 1921. A minimum average life of 1000 hours for a sample of not less than 5 lamps selected at random from a bulk supply was specified. (This followed the practice which had been accepted for over thirty years for carbon lamps, the average lives of which were specified as 400 hours and 800 hours for two types of lamp.) The manufacturer has to adopt an objective average life greater than 1000 hours.

of the lamps which would fail before some defined number of hours to be readily estimated. It became general at that time to consider the reduction in the proportion of failures before 500 hours to be the main technical objective. The summary report for the year 1924 contained histograms and survivor curves for every watt rating which represented

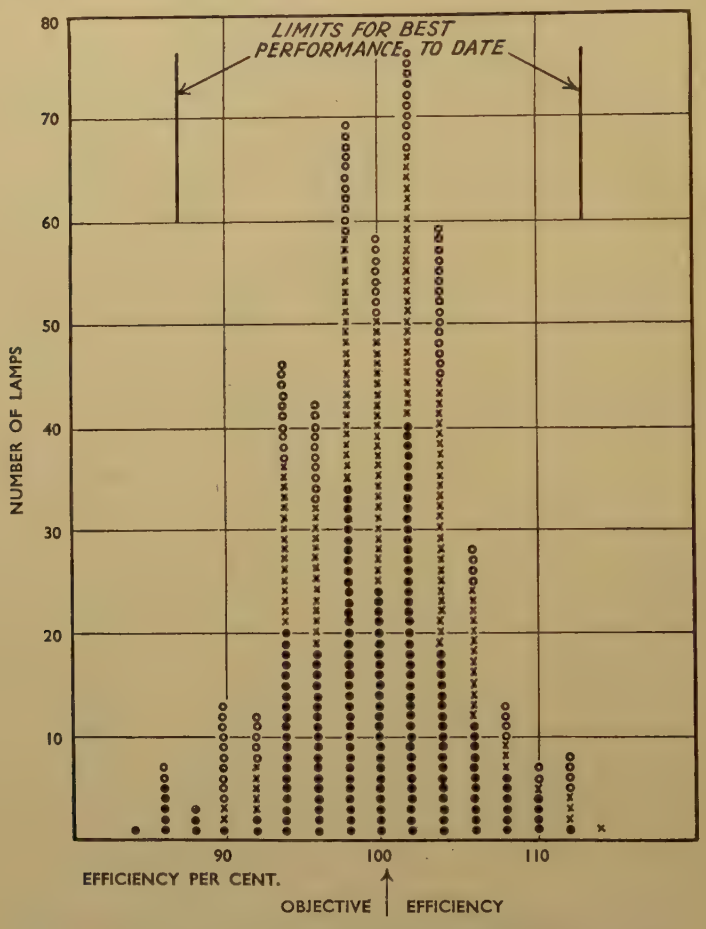


Fig. 1. Frequency distribution of electric lamps classified according to efficiency. The lamps were collected during 1924 and those collected in different months are represented by different symbols.

more than 3 per cent. of the total production of lamps used for general lighting. The performance of lamps of these ratings was the subject of the earliest specification for tungsten filament lamps issued by the British Engineering Standards Association (later the British Standards Institution) in June 1921.

The following is a quotation from a report issued to the factory in July 1925 immediately after the circulation of the annual report: 'One of the main objects (of this report) is to determine whether a definite criterion can be laid down for indicating whether or not lamps are of

the best quality and to devise a method by which the works could be advised of fluctuations in quality more readily than is possible by the mere inspection of life test sheets.'

The first scheme introduced to the factory involved selecting five lamps at random from the lamps manufactured each half-day or each day. The initial efficiencies of these lamps were plotted in the form of a histogram, the results for each watt rating being shown on a separate diagram. As the designed or objective efficiency of lamps in each watt rating differed for each voltage rating (200, 210, 220, 230, 240, 250, 260) all results were expressed as a percentage of the objective values. The results for lamps of different voltage rating were indicated by a different symbol, and the results for one month were shown on one histogram on which limit lines were also drawn within which all results should lie. These limit lines were originally based on the best results previously achieved, thereby creating a technical incentive to improve. They were reviewed and the tolerance reduced as improvements were achieved (see Fig. 1).

In addition, separate histograms were maintained for each voltage-wattage rating, but owing to the inevitable differences in the demand for the various voltages these diagrams would include the data for many months, and the results for each month were indicated by a separate symbol. These diagrams served to illustrate clearly whether the results for any particular voltage rating in a watt rating showed abnormal variations or were consistently above or below objective. As early as 1920 monthly meetings were arranged at which laboratory staff and works technical supervisory staff reviewed the work in hand. The diagrams above referred to were considered at these meetings in 1925 as well as being scrutinised daily by the works staff.

These diagrams and similar diagrams for other routine data were, in fact, control charts, although they were not so called. However, their use and acceptance as important contributions to help supervision paved the way for the introduction of the Shewhart type of control chart a little later.

Experimental Investigations

Laboratory reports of 1924 and 1925 show that considerable laboratory effort was devoted to reducing the variability found in the initial efficiency of lamps. The reduction of this variability was important because the lives of lamps change inversely with the 6th to 7th power of the initial efficiency. While the scientific staff were using the standard deviation and the coefficient of variation as measures of dispersion and were utilising the method of analysis of variance when studying test data, these statistical terms were not introduced to the factory personnel until about a year later. Coefficient of variation was a particularly convenient method of expressing the dispersion in the initial efficiency of lamps because the objective efficiency of a lamp increased with the diameter of the filament wire, and hence, for a

given degree of dispersion, the magnitude of the standard deviation of efficiency also increased when expressed in photometric units; but the coefficient of variation was practically constant for lamps varying in watt rating from 20 watts to 1500 watts. It is convenient to note here that the coefficient of variation in efficiency was specified first in the

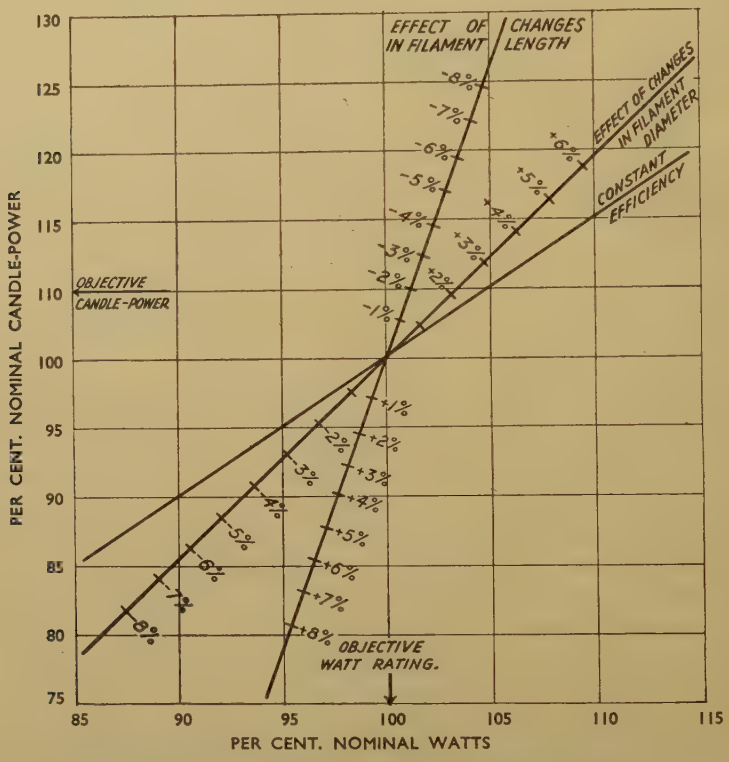


Fig. 2. Relations between candle-power and current of incandescent lamps for small changes in filament diameter and filament length. (Straight-wire lamps at constant voltage.)

British Standard Specification issued in 1932, and I believe this was the first national specification in the world to contain a clause placing a definite limit of this kind on the variability allowed in a characteristic of a manufactured product.

Arising out of the experimental work already referred to, the diagram shown in Fig. 2 was prepared in 1924 to show how filament dimensions affected the initial rating of vacuum lamps.[5] About a year later a similar diagram was prepared for the gasfilled lamps, whose rating is not only affected by the diameter and length of the filament wire but also by the diameter of the wire on which the filament wire is coiled and the distance between the turns of the coil.[6] Errors in this last dimension can arise in the coiling operation, during heat-treatment, and during the mounting operation. The diameters of the wires concerned in the lamps most commonly used are from 0·0005 in. to about 0·005

in.; the errors in filament diameter which are of technical importance are of the order of 0·00001 in.

In addition to experimental work under laboratory conditions experiments were carried out under the normal manufacturing conditions. A characteristic of much of this experimental work was the determination of the degree of variability in the final product of a series of processes, contributed by the various stages in manufacture. Experiments in glassmaking, metal-making, and lamp assembly were devised on the following general principles.

Material from one department, usually handled as a batch, was supplied to different operators or different semi-automatic or automatic machines in the next department, and a comparison made for different operators and machines in that department of the variability in the quality of the final product or in the amount of material wasted. Or, a batch of material was put on one side and a day's supply from this batch issued to an operator or a machine at intervals of a week, thereby making it possible to estimate the variability introduced by using the same class of material at different times. In some experiments material or components were graded into two or more groups and the performance of operators or machines, supplied with the different grades of material, compared. It is obvious that by such means the relative importance of various factors affecting production could be estimated. It is true that some of these experiments could have been combined and the results subsequently analysed, but this procedure was rarely followed because increasing the size of the experiment or the time occupied in carrying it through increased the possibility that some other factors not under investigation would change and complicate the interpretation of the results. Furthermore, the simpler type of experiment provided a better means of interesting factory technicians in the principles which governed the analysis of the results.

An example of this relatively elementary type of investigation is now briefly described because during recent years I have promoted similar simple investigations in industries concerned with the manufacture of building bricks, pottery, foodstuffs, and clothing, which have yielded valuable results. These later investigations have always been associated with a use of the now well-known control chart technique.

Reference has already been made to the effect of errors in the dimensions of coiled filaments on lamp rating. It was found that stretching a filament during the process of mounting it on the supports and leading-in wires could have a major effect on the efficiency of the finished lamp. This process was carried out by hand, using special tools, and about a dozen operators were needed to produce lamps of each of the ratings in most common demand. For the purpose of the experiment, a batch of about 3000 filaments was set aside.

A batch of filaments was made from one length of tungsten filament wire, wound into a continuous coil on a length of iron wire, the length of coil subjected to heat-treatment, the coil cut up into pieces of

specified length, and finally the iron mandrel removed from all the filaments by a series of chemical treatments. The experiment was carried out as follows:

(i) About fifty filaments were mounted by an operator on each of several occasions, and the glass mounts marked for identification, the mounted filaments being made up into lamps on each occasion; the lamps were evacuated in order to avoid deterioration of the filament. Later all these lamps were mixed together, baked, re-evacuated, filled with gas, and completed, as a batch. All lamps were measured by the same operator.

(ii) About fifty filaments were issued to each of fifteen operators on the same occasion, the mounts marked for identification, and all the lamps mixed for completion of manufacture and for measurement as in (i).

The results were presented to factory personnel in the form of histograms similar to those illustrated in Fig. 1. These histograms demonstrated that an operator introduced additional variability when mounting filaments on different occasions and that even greater differences were introduced by the different operators on the same occasion. By comparing the results with the histogram obtained when plotting the routine test results obtained in six months, it was obvious that differences in mounting techniques were a major contribution to variations in the initial rating test results. To reduce the variability introduced by filament mounting an automatic machine was ultimately designed which was steadily improved by the engineers, improvement being measured by a reduction in the variability in the efficiency of lamps containing filaments from the same batch.

Investigations of a similar kind were carried out on glass bulb and tube-making processes, on chemical and mechanical processes used in the production of filament wire, mandrel wire, and leading-in wires, and on the operation of many semi-automatic and automatic machines. The systematic study of the sources of variability made all supervisory staff more conscious of the importance of obtaining measurements of variability, and it soon became commonplace to use evidence of consistency in the standard deviation of small groups of test results as one measure of the quality of a process.

The recording and presentation of data collected as a routine at all stages of manufacture was under constant review, and the publication of Dr. Shewhart's paper on Quality Control Charts in 1926[7] led to an important and extremely valuable development. The practice of plotting on control charts the average and standard deviation of small groups of observations as a regular routine soon displaced other methods of presentation. Factory personnel were encouraged to indicate on the charts events which occurred in the factory, such as change of operator, breakdown of a machine, introduction of a new batch of material, etc. Such notes are of great importance when reviewing control charts accumulated over long periods. Further, besides indicating on the

charts the limits recommended by Dr. Shewhart, within which all results should lie, additional limit lines were drawn within which about 95 per cent. of the results should lie. This was done to help the factory staff to appreciate more readily how the plotted points were distributed within the limit lines. It also tended to call attention to the initiation of any gradual changes in the position of the plotted points, sufficiently early to enable effective technical action to be taken. In this way the charts served in helping to arrest changes before they became economically serious. If these 95 per cent. limits were not used there was a tendency for recognition of such changes to be delayed so long that production difficulties became serious.

The use of the range in small samples of measurements as a measure of variability was not introduced till the later twenties, and even then for some time both the range and the standard deviation were determined and plotted on charts. By 1930 these charts were regularly maintained at about twenty of the important stages in glass and filament metal manufacture and in lamp assembly. Studies and analyses of control chart data were used to indicate profitable lines of development and research.

It is not possible to continue to describe in detail the gradual evolution of control chart techniques suited to different types of process and testing procedures. Many of these are described in the pamphlet issued by the Laboratories during the recent war.[8]

GENERAL COMMENTS

A very important feature of the changes in mode of thought promoted by the use of this statistical technique was the concentration of everybody on the performance of machines or processes rather than on the final product. It is desirable to state more clearly what is meant by concentrating attention on processes.

Inspection of final products has been long established as a means of determining whether or not a product satisfies a specification. Similar procedures have been adopted in factories and have become traditional. The method quite commonly used involves setting up an inspection department completely divorced from the production department. In these circumstances the results of inspection are often not available to those concerned with supervising the production processes for several days. (One of the production managers whom I helped to install control chart techniques summarised this position aptly by a notice he had placed in the inspection department: 'Our results are too comprehensive and too late.')

Another procedure commonly adopted when manufacturing articles to dimensional limits is to use mobile inspectors who visit machines at intervals, testing or measuring a few articles and making adjustments. While this method has much to commend it, the principal criticism of it is that it takes no cognizance of the magnitude of the variability

which is inherent in the product of the machine even when operating in a manner judged to be satisfactory on general technical grounds. A further disadvantage is that no record is kept of the frequency and nature of the adjustments made or of their relation to other changes which are inevitable under production conditions. If this form of inspection be applied to the product of processes which are more difficult to control and have greater inherent variability, then the adjustments made as a result of the mobile inspection can lead to an increase in the variability of the measured characteristic of the product. This is due to the fact that fairly large deviations from the objective arise solely by chance from the operation of factors which are always present in such a process, and any 'correction' made on the basis of these chance results unnecessarily disturbs the process. When this condition exists the introduction of control charts on which results of periodical inspection are plotted leads to much less frequent and less drastic adjustment of the machine or process and to an improvement in the uniformity of the product; economies are made in inspection and in any subsequent processes because of the improvement of the uniformity of the product supplied. In one instance, where the introduction of control charts had greatly reduced the apparent activity of the machine minder, I suggested to a senior manager that he should not visit the machine too frequently, as the operator immediately began to use the many adjustments provided, thereby upsetting the process, which had subsequently to be readjusted.

Type and Accuracy of Measurements

When a system of inspection and presentation of results of the kind just described has been in regular use for some time, those concerned begin to realise that they concentrate much more on the condition of the process than on the final product of the factory. If some considerable amount of final inspection of the product has continued, which is usual while the control charts are being introduced, the staff find by experience that faulty product of the type and degree previously found is rarely discovered. It is then a relatively easy step for the technicians concerned to realise that in some cases it is more direct and more appropriate to record regularly some feature of the process which is generally accepted as important. This may involve introducing some simple measuring devices designed specifically to meet the practical conditions.

If measurements are made on the characteristics of processes, machines, or products, it should be emphasised that in the design of the measuring devices it is not necessary to aim at the repetitive accuracy that might be deemed necessary in a laboratory, and that the most important thing is to avoid bias creeping into the measurement. The accuracy required depends on the degree of variability that is tolerable in the process characteristic being examined. Technicians who have had explained to them the elementary principles which

govern the position of the limits on the control chart soon appreciate that the standard deviation of repeated measurements on precisely the same material or article can be as large as from one-third to one-half the standard deviation which is inherent in the product when the process is operating satisfactorily. That is, they learn that the apparent standard deviation in the characteristic under observation is only 10 per cent. larger than the real variability which would be determined if there was no contribution whatever from the variability in the measurements.

Further, to avoid any bias affecting the measurements every effort should be made to use substitution methods of measurement. The principle involved is to use calibrated substandards similar to the material or article measured, thereby reducing the effect of zero scale errors in the measuring devices. By using substitution methods of measurement it is possible to arrange that any indicating instrument used reads positive differences from the standard in suitable units of measurement. This reduces the results to quite small numbers, thereby simplifying the arithmetical work and tending to the elimination of errors in calculation.

Details concerning the Introduction of Control Chart Techniques

There are a number of details, some, if not all, of which arise when the introduction of control chart technique to a factory organisation is discussed.

The term 'statistics' suggests to the minds of many people the collection of masses of figures in tabular form, and they will frequently claim that they are already using statistics to guide technical control, and the mention of quality control will provoke the comment that the procedures adopted for carrying on production are largely concerned with the control of quality. Although it must be admitted that the duties of many technicians in industry are directed towards controlling or even improving the quality of their product, some criticism of the more usual methods has already been made. The advantages of sample inspection at key points in the production line, coupled with pictorial presentation of the results in the form of a chart on which limits are indicated, cannot be established by argument, but rarely is it difficult to establish the advantages this procedure has over traditional inspection once the routine is in operation on a reasonable scale.

Sometimes 'statistics' is considered synonymous with advanced mathematics, and many of the proposed curricula issued by the mathematical departments of technical colleges give ample grounds for this view. This erroneous conception can only be dispelled by actual demonstration in the factory of the procedures necessary for applying the control chart technique.

Type of Personnel. In practice the construction of a control chart calls for nothing but the simplest arithmetic, and reference to tables less voluminous than many which are in common use in factories. The

work can be carried out satisfactorily by persons of average intelligence who have only had elementary school education, although there are some advantages in using persons who have attained the level demanded by the school certificate. In the majority of cases the person constructing the charts can do the sampling and make the measurements required.

It is appropriate at this stage to note that in the older forms of routine inspection those regularly engaged on this work have a relatively dull job which can be very fatiguing, and this can be a source of errors. Sampling inspection at different points in the production line adds considerable interest to the work, and this interest is augmented by having to construct the chart. The change of duties tends to reduce fatigue, and this coupled with the interest stimulated leads to a high standard of performance by the person concerned.

Intermittent Production or Short Runs. Because quality control was introduced during the war to organisations concerned with the quantity production of munition components many people formed the impression that it was suited only to this type of manufacture. There is no doubt that the application of control chart technique can be most readily applied to quantity production of this kind of goods. When this is admitted, factory personnel will try to excuse themselves the initial effort required in applying the technique to their particular work and on some occasions emphasise how frequently there are changes in the article they have to make. Earlier it has been pointed out that the technique really concentrates attention on the process and that the control chart is a pictorial method of indicating whether or not the process can fulfil the duties it has been designed to do. The nature of the operations carried out on a machine or process is not usually greatly modified because the article produced is somewhat different in size and shape. Again, the persons concerned cannot be convinced of the correctness of this contention, but it is quickly appreciated in practice after the charts have been introduced.

How and Where should the Technique be Introduced? It is likely that the reader will have noticed the continued reference to technicians in the foregoing and how rarely executive management is mentioned. In practice the introduction of a novel technique to a factory is best done at those levels of personnel directly interested, and all that is asked of senior management is encouragement and confidence. Even if senior members of management were convinced of the utility of statistical methods as everyday tools of production it would be natural for a manager to expect a reasonably cut and dried scheme, with estimates of cost and of potential savings, before he sanctioned actual work. He would also expect to initiate the introduction of such a new technique on a definite date. Because the modes of thought are somewhat revolutionary it is a practical necessity to introduce the use of the techniques which form the subject of this article in a somewhat experimental way. Additional reasons for doing so are that it is imperative to utilise fully

the technical knowledge available in the organisation and to adjust procedure to suit the existing staff arrangements. In any reasonably well-run production unit it is also unlikely that the early exploratory work will necessarily prove to be along the most suitable lines. Even when collaborating with managements desiring to introduce the techniques, my colleagues and I have always urged the desirability of commencing on a small scale, choosing some item or part of the manufacturing process after discussing with the appropriate members of the staff the nature of their most common difficulties and the technical background which governs their activities.

Unless the initial approach came from a senior member of the management, the policy I have adopted on all occasions is to interest some relatively junior supervisory technician to explore the new technique on the grounds that it will help him in his everyday duties, offering to give him advice after he has collected data for, say, two to four weeks. This method of approach has met with success in a very high proportion of the cases in which it has been tried, and a little success will always gain the support of more senior management much more readily than any argument, however carefully staged. There have been many occasions where the success has been such that it is not necessary to emphasise the importance of extending the use of the technique.

Much of the foregoing commentary on the introduction of the control chart technique to a factory organisation can be summarised by stressing the fact that it is essentially a technique which is complementary to technical knowledge and experience, and not a substitute for them. All techniques, whether they be concerned with playing a musical instrument or playing a ball game, demand practice by those aspiring to achieve success. Techniques cannot be acquired by attending lectures, listening to learned discussions, or by reading books. Given perseverance to practise the technique coupled with an average degree of intelligence, any person can make rapid progress; and this can be assisted by attending lectures, etc. Successful application of a statistical technique to improve efficiency of production also demands a good knowledge of the sciences and technologies concerned. The better the knowledge and the higher the intelligence of the person overseeing the work the more rapid will be the progress and the fewer will be the false steps made. This is obviously a truism and applicable to all kinds of human activities.

It cannot be over-emphasised that the use of statistical methods will not solve problems demanding technical or scientific knowledge, but it will frequently emphasise lack of these attributes and stimulate searching enquiry where, without this stimulus, an atmosphere of false satisfaction may be created only to be destroyed by some catastrophic experience, more expensive and more painful than any likely to be encountered in introducing statistical modes of thought to guide everyday decisions.

Personal Relationships

If a trained statistician is concerned with giving guidance to factory personnel initiating the use of control chart technique he should be guided by the following general considerations.

Those responsible for setting up industrial processes design machines and adopt procedures based on very considerable knowledge and experience, and in general have a good appreciation of the more important requirements that the processes and machines should meet. While it is true that often it will be found that the machines and processes do not conform consistently to the objective as well as the designers expect, it is very important that the statistical adviser (or any trained scientific specialist) should approach his work with an attitude of mind which gives due credit to the experience and knowledge of those he is trying to help. He should avoid any appearance of conceit in his own powers or in the importance of his specialised training. By so doing he will not find it difficult to gain the good will of those with whom he is collaborating, without which failure to give satisfactory service is inevitable. Frequently the practice of control chart techniques does expose weaknesses in technical knowledge or incorrect association of cause and effect, but this will cause no heartburning and give rise to no difficulties if the statistical adviser has convinced those charged with the duties of supervising production plant of his sincerity to act as a sympathetic helper.

Control Chart Techniques and Industrial Efficiency

While it is not possible in an article of reasonable length to enumerate and discuss all aspects of the assistance which the use of statistical methods can give to a manufacturing organisation, the following brief summary may be of interest.

When control charts are accepted as the medium for presenting the results of inspection at the key stages of production *and are regularly studied* they serve three main purposes:

(i) *They assist the planning and supervision of production.* The charts present in a concise form a long-term yet up-to-date picture of events and should be used to assist day-to-day technical supervision of production. They also provide a measure of what can be expected from operators, machines, and processes, which is useful when planning production programmes.

(ii) *They improve personal relationships.* My experience in very different types of industry and with different production methods, extending from skilled craftsmanship to fully automatic machines, leads me to think that it is in the field of personal relationships that quality control charts make their most important contribution. They foster interest in tasks which often contain little to interest the person carrying them out; they improve personal relationships between the relatively junior staff in charge of different sections of

a manufacturing organisation and in particular between inspectors and technical supervisors; they provide a common method for presenting and studying data on a long-term basis which can be readily understood by all concerned whatever their function or rank; and discussion is guided by reason supported by fact rather than by traditions and hunches.

(iii) *They guide development and research.* The data recorded on control charts can be used for formal statistical investigations when considering long-term projects aimed at improving the quality and reducing the cost of the product. The planning of statistical investigations of this type and their interpretation demand a detailed knowledge of the scientific and technical background of the industry which is not normally possessed by persons capable of carrying out the investigations, but those engaged in the industry who possess such knowledge can greatly assist the formal investigations if they have some knowledge of statistical principles and have acquired the habit of thinking statistically. By this combination of technical knowledge and statistical thought development is more likely to be planned along lines which will lead to the desired improvement at the lowest cost.

REFERENCES AND BIBLIOGRAPHY

[1] KENDALL, M. G. (1943). *The Advanced Theory of Statistics.* Griffin, London.

[2] SHEWHART, W. A. (1924). 'Some Applications of Statistical Methods to the Analysis of Physical and Engineering Data,' *Bell System Tech. J.*, **13**, 43.

[3] MOLINA, E. C., and CROWELL, R. P. (1924). 'Deviation of Random Samples from Average Conditions and Significance to Traffic Men,' *ibid.*, **3**, 88.

[4] PATERSON, C. C., and DUDDING, B. P. (1915). 'The Unit of Candle-power in White Light,' *Proc. Phys. Soc. Lond.*, **27**, 263.

[5] DUDDING, B. P., and WINCH, G. T. (1925–6). 'The Characteristics of Electric Lamps in Relation to their Testing,' *Proc. Inst. Engng. Insp.*, p. 35.

[6] CHELIOTI, G., and DUDDING, B. P. (1931). 'Precision in Incandescent Lamp Manufacture,' *G.E.C. J.*, **2**, 103, and *Proc. Int. Congr. Illum.*

[7] SHEWHART, W. A. (1926). 'Quality Control Charts,' *Bell System Tech. J.*, **5**, 593.

[8] DUDDING, B. P., and JENNETT, W. J. (1944). *Quality Control Chart Technique when Manufacturing to a Specification*, Research Laboratories, The General Electric Co. Ltd., Wembley, Middlesex.

The following are papers by the author and his colleagues, additional to those directly referred to in the article.

DUDDING, B. P. (1928–9). 'Errors in testing Bulk Supply by Random Selection,' *Proc. Inst. Engng. Insp.*, p. 61.

DUDDING, B. P., and BAKER, I. M. (1933). 'The Application of Statistical Methods to the Quality Control of Manufactured Products,' *Trans. Soc. Glass Technol.*, **17**.

Idem. (1934). 'The Application of Statistical Methods to the Planning of Routine Testing Procedure,' *ibid.*, **18**.

JENNETT, W. J., and DUDDING, B. P. (1936). 'The Application of Statistical Principles to an Industrial Problem,' *Suppl. J. R. Statist. Soc.*, **3**, No. 1.

DUDDING, B. P. (1936). 'The Application of Statistical Principles to Industrial Problems,' *Jun. Instn. Engrs.*, **46**, 278.

PATERSON, C. C. (1936). 'Uniformity as the Gauge of Quality,' *J. Instn. Elec. Engrs.*, **79**, No. 480.

DUDDING, B. P. (1946). 'A Further Study of the Performance of Glass-melting Tank Furnace Fired by Coke Oven Gas,' *Trans. Soc. Glass Technol.*, **30**, 13–30.

DUDDING, B. P., and JENNETT, W. J. (1937). 'Statistical Technique as a Tool in Industry,' *G.E.C. J.*, **8**, No. 3.

Idem. (1940). 'Statistics and Engineering Practice,' *J. Instn. Elec. Engrs.*, **87**, No. 523.

DUDDING, B. P. (1942). 'Sampling Inspection and Quality Control,' *J. Instn. Prod. Engrs.*, **21**, 13.

Idem. (1942). 'Statistical Methods as Aids to Production Efficiency,' *Manchr. Statist. Soc.*, April 1942.

DUDDING, B. P., and JENNETT, W. J. (1942). 'Quality Control Charts,' *B.S. Instn. publication* 600*R*.

Idem. (1945). 'Quality Control—Use of Statistical Methods in Metallurgical Industry,' *Metal Ind.*, March 1945.

DUDDING, B. P., and KEEN, JOAN. (1949). 'The Use of Observations on Samples as an Aid to Production Efficiency,' *Nat. Hosiery Mfrs. Fed. Conf.*, Harrogate.

Idem. (1949). 'Statistical Methods and Specifications,' *International Statist. Inst.*, 26th session.

DUDDING, B. P. (1950). 'A Survey of Present Methods of Industrial Statistics,' *J. Instn. Engng. Insp.*, **14**, 78.

A STATISTICIAN LOOKS AT THE INTERNATIONAL MANAGEMENT CONGRESS OF 1951

ALBERT W. SWAN, B.A.Sc., A.M.I.Mech.E.
In charge of Operational Research, Courtaulds Ltd.

As Mr. Swan was going to the Brussels conference he was invited to write his impressions from the point of view of the applied statistician.

Mr. Swan says that a certain amount of time was devoted to the use of statistical thinking as a tool for management, and infers from the interest expressed that more attention will have to be given to statistics at the next conference. This time statistics entered into the discussions of work measurement, quality control, and job evaluation.

Introduction

The conferences of the Congrès Internationale d'Organisations Scientifiques (CIOS) have been held to assemble current views on various aspects of management at 3–4 year intervals except during the war. These notes are based on impressions of the congress held at Brussels in July 1951.

The congress lasted a week, the working days consisting of formal meetings from 9.30 to 12 a.m. and from 2.30 to 5 p.m., with various informal meetings outside these hours. For the Brussels congress a careful preliminary concentration of the material available was done by international panels, and this was followed by the publication of papers more or less representing the views of the panels. These papers were circulated to delegates before the congress, and the meetings took the form of a spoken summary of the published report followed by a discussion, usually in the form of written questions submitted by delegates and dealt with by one or more members of the panel. This general method had two disadvantages: the reports published by the panel tended to be rather colourless as endeavouring to reconcile opinions which often differed greatly, and the 'discussions' were almost always too much in the hands of the panel members on the daïs. It would, however, be difficult to work out a practical alternative; the chairmen were able and fair, and the discussions were usually sensible and added to points brought out in the reports.

All meetings were conducted in French and English as the official languages, with interpreters to provide running translations. These translations were transmitted on the spot by microwave wireless and heard on portable wireless sets with earphones, each set being adjusted by the delegate to the language desired.

Before proceeding to comment on the meetings which the writer attended it should be pointed out that the congress was not concerned

with statistics as such, but a certain amount of time was devoted to the use of statistical thinking as a tool for management, and it was clear from the interest expressed that a larger proportion of the next meeting in 1954 will perforce be given to statistical matters. The comments which follow include summaries of those meetings attended, plus some of the writer's suggestions not stated at the time of the meeting. All opinions expressed are personal.

Work Measurement

This was a disappointing meeting to a statistician. Well-deserved praise was given to the work done by Mr. D. J. Desmond on the method of analysis of the variability of time-study 'rating,' and it was made clear that work is being done by Professor Mundel of Purdue University and others in the USA, and by Professor Matthew of Birmingham University and others in Great Britain, in an attempt to find a statistical basis for setting up standards for rating.

The rating method of making a time study is to time the operation by a stop-watch and then to apply a rating, the value of which depends upon the observer's judgment as to how hard the operative is working. No reference was made by the American or British members of the panel to the fact that there is a school of thought in both the USA and Great Britain which regards as suspect the taking of an objective time study and then altering the time obtained by applying a subjective judgment.

Readers may be interested in a notable study of the variability of observers' subjective judgments on rating, *The Accuracy of Time Studies*, by Mr. W. Kinniburgh and Mr. L. S. Vallance of the Department of Scientific and Industrial Research Building Research Station. The statistical approach to time study in the USA is illustrated in Dr. Wm. Gomberg's *A Trade Union Analysis of Time Study*, Mr. L. A. Sylvester's *The Handbook of Advanced Time and Motion Study*, and various papers by Prof. S. B. Littauer of Columbia University.

At the congress a considerable amount of attention was devoted to a recent development known as 'methods-time-measurement,' which purports to build up total times for jobs from predetermined standards for elementary movements. The notion of variability does not appear to be included in the scheme.

The statistical aspect of time study had its turn when the French and Belgian speakers had their say. Useful investigations have been made, particularly in France, and it was interesting to note a tendency to use the modal or most common value of the times taken rather than the average. This was presumably because distributions of times are by no means necessarily normal. The writer has found that the shape of distribution curves may be of extreme importance in connection with time studies and that the types found may be classified as follows:

1. Single peak, normal type, narrow—signifying uniform practice: good.

2. Single peak, normal type, wide—signifying uniform practice: not so good.
3. Two-peaked—signifying two sets of conditions such as the operation of learners and experienced workers.
4. Skew, tail towards longer time—signifying mostly uniform practice with some irregularity.
5. Rectangular—signifying chaotic conditions, sometimes pointing to trouble at previous operation.

Recent Developments in Quality Control

This was an interesting and lively meeting with plenty of discussion. The basic report by Mr. D. J. Desmond was a survey of the field of quality control, giving a number of recent developments in Great Britain and elsewhere, and containing a useful section on recent methods of simplification. The discussion centred on three main topics: (1) Whether or not the use of quality control has declined in Great Britain and, if it has, why it has so declined; (2) Whether or not the term 'quality control' is too limited in view of the current expansion in the use of industrial statistics; and (3) The limitations of quality control in relation to 100 per cent. inspection.

Mr. Desmond stated categorically in his report that quality control suffered a setback in Great Britain after the war, and it was generally agreed that within the definition of quality control as agreed by the panel (in effect the use of control charts and statistical sampling plans) this was true. A reason advanced by one member of the audience was that while quality control would apply well and economically where the scrap rate was high, say up to 10–15 per cent., as was true for certain products during the war, it might prove too expensive where the usual scrap rate is 0·5 or 1 per cent. There is a simple answer to this point, but it was not put forward at the meeting. The subject 'The Economic Limitations of Statistical Quality Control' might be suitable for one of the Royal Statistical Society Industrial Applications group meetings.

Several speakers agreed that the term 'quality control' is now too narrow and should be widened to some such term as 'statistics in industry' or 'operational research,' since the same basic method of thinking is used in applying a wide variety of statistical methods to a still wider variety of industrial problems. The writer would point out, however, that the term 'operational research' is still far from having a single definition. There are those who believe that the only difference between operational research and scientific management, a term introduced thirty or more years ago by Taylor, Gantt and Gilbreth, is that the former includes the statistical approach. Others agree with this in general, but consider that operational research leads to, but does not include, routine control and therefore should not include much of the work done as statistical quality control.

One speaker at the congress stated that if the wider view is taken of the position of statistics applied to industrial problems rather than of

quality control the picture of activity in Great Britain is much more pleasant to contemplate, that there is in fact a rapidly expanding use of statistical thinking. The writer is aware that three of the largest groups of works in the British steel industry have recently followed the example of another group by starting statistical departments for investigation and control. Among other instances of statistical applications in industry in Great Britain known to the writer are the studies of airport delay times by Dr. Bell and his colleagues of the Ministry of Civil Aviation, the Operational Research and Statistical Departments in Imperial Chemical Industries Ltd., Shell-Mex Ltd., English Electric Co. Ltd., and British Nylon Spinners Ltd. This list is far from complete.

The topic 'How can quality control be considered in relation to 100 per cent. inspection?' was dealt with from the audience by Mr. Loxham of the Sigma Instrument Company. It should be stated that 100 per cent. inspection (inspection of all units manufactured) is common in certain types of industry, particularly engineering, whereas statistical quality control is essentially based on sampling. Mr. Loxham's first point was that in any engineering works in which parts are made to extreme precision and where a single failure may be catastrophic, as in the case of blading for gas turbines, 100 per cent. inspection cannot be avoided. However, any part such as a turbine blade has a number of dimensions to be checked, and quality control should be, and in his experience is, used at an early stage in manufacture to determine the variability for each dimension and to direct attention to the critical dimensions. His second point was that mechanical and electronic 100 per cent. gauging have now been developed to the extent that components can be checked 100 per cent. for several dimensions at a very high speed and therefore at low cost, with an overall error of one-tenth of the tolerance. Where this is the case 100 per cent. gauging should be used and not quality control.

Working Methods for Top Managers and Education for Management

Neither of the meetings on these two subjects was concerned with statistics in industry as such, but there was included in the panel reports and discussions much that is of practical interest to the industrial statistician. In the writer's view the psychological aspect of the application of statistical methods in industry is one which has had little formal attention, yet on the knowledge of how managements think and behave and on the use of this knowledge depend the success or failure of a particular application. It is highly important to know and understand the management structure of the organisation in which one works, how decisions are formulated (as for instance by committees or individuals) and how they are put into action.

Perhaps the most interesting point that emerged from the two meetings was the general recognition that management is itself a technique and that the skill it requires is quite distinct from the skill

required for the process controlled. The head of an industrial statistical department, for example, must have or acquire skill in managing.

Free evening classes in management have been organised in several centres in France in order to help smaller companies who feel unable to spare promising members of their staff to attend full-time courses. These classes last for periods from a week up to several months and are a notable success.

An interesting sequel to the public meeting was a meeting under the auspices of the panel of the professors of engineering production, industrial engineering and administrative engineering in attendance from Great Britain, Canada, the United States, Switzerland, Sweden and Holland. As a result it was decided to reinstate an international committee of professors for the purpose of corresponding on teaching methods and on the compilation of uniform syllabuses.

Job Evaluation

The term 'job evaluation' is used to describe the attempt which is now being made to rate the relative difficulty and importance of jobs so that pay can be established on an equitable basis.

One school of thought prefers to have relatively few classifications, say three or four, on the grounds that if there are too many there are certain to be inter-correlations that are difficult to estimate, and also that the complete set of classifications may be difficult to explain. Another school of thought prefers more detailed sub-classifications. But it is generally agreed that there must be at least three, relating to the function, the worker, and the external circumstances. A simple twofold classification on this basis might be to decide that the work is easy or difficult, the worker is young or old, and the work clean or dirty; a more complex sub-classification would be to provide a numerical score from say 0 to 5 for each main category, giving say 3 under 'difficulty' if the work is pretty difficult, 2 under 'age' if the worker is between 20 and 25, and 4 under 'conditions' if the work is core-making in a foundry. The inter-correlation can be illustrated from this hypothetical case; the young worker might find any given job easier than an older worker, and be more affected by unpleasant working conditions.

The most interesting and valuable feature of this meeting to a statistician was the surprise announcement of the completion under the auspices of the CIOS Panel by Professor Matthew of Birmingham of a statistical study comprising an international comparison of job evaluation. The report of this study will be available, but in the meantime some of the main features may be mentioned.

The report is entitled 'An Experimental Approach to determine the Extent to which Job Evaluation carried out in Different Countries using "Points" Schemes are Comparable.' The investigation dealt with seventeen jobs sent in by groups from eight countries (five from France), and its objects were to determine differences between groups

and between jobs, the range of variation inherent in job evaluation, and the extent of the usefulness of international comparisons. The main results were that after certain precautions had been taken in regard to the differing scale values used, the difference between the means and variances of ten out of twelve groups and between all of the seventeen jobs were not significant. These are important results, since they show that international comparisons are possible, and that comparisons among jobs (machinist, craneman, etc.) are also possible.

The writer is informed that at a panel meeting following the public discussion it was decided to set up an international committee which would carry forward during the next two or three years the international comparison of job evaluation on a statistical basis, continuing the work initiated by Professor Matthew.

Meeting of the Société Belge de Statistique

An interesting informal meeting was that of the Société Belge de Statistique to which Mr. Desmond and the writer were invited as representing the Royal Statistical Society. There was no programme but there were a number of interesting private and semi-private discussions, leading in several cases to invitations to visit works. The writer, for instance, accepted invitations for discussions with representatives of Fabelta (Belgian rayon manufacturers) and the Union Cotonnière. Much active work in statistical methods is being carried on in the Belgian textile industry, work which should be better known in Great Britain.

Conclusion

The conference was serious in purpose and in act; continuous effort was successfully made to keep away from high-sounding generalities and to be practical. The discussions added considerably to the material in the reports, and the informal meetings were most valuable. It was evident that the statistical way of thinking is making rapid headway in higher management.

ESTIMATING THE SOCIAL CLASS OF TOWNS

LESLIE T. WILKINS
The Social Survey

A measure of the social class of towns that is useful to businessmen and economists is the percentage of households in which the chief wage-earner has a basic wage exceeding £7 10s. a week. Data are not available for determining this directly, but this article shows how the measure may be estimated from available data and gives a table of values for 140 large towns.

Most people assume that they know what is meant by 'social class,' and it is also generally assumed that towns and other communities possess characteristics which determine their social status. For example, areas which are populated in the main by 'working-class families' are referred to as 'working-class districts.'

Nevertheless 'social class' has, in common with 'intelligence,' the peculiarity that while the man in the street is quite happy with the term, experts fail to agree. Indeed, it is a concept that has quite a lot of technical similarities to the concept of intelligence. We shall not go into these matters in any detail but concentrate upon a practical application of the concept of the social class of areas, which may be of use to businessmen and social scientists.

About a year ago the author was asked by the editor of *The Marketing Survey of the United Kingdom*[1] to suggest a method of constructing an 'income levels index' for the larger towns. This article reports a slightly modified form of the method used in the earlier work; the modifications make it more rigorous and possibly more useful to social scientists without reducing its value to businessmen.

An Income Measure of Social Class

There is, of course, no method of measuring the social class of towns, even if it were possible to define what was meant by the term. We can, perhaps, best describe a town's status with reference to the people who live in it, selecting some measure of the living standards of these persons. One frequently used measure of these standards of living is the income of the heads of households. In practice, again, it is impossible to obtain information regarding income distributions, other than the national figures published by the Inland Revenue. These figures are not broken down by areas or towns; only the global figures are given, and even these have many severe limitations for social research purposes.

It is, however, customary for persons interviewed in the course of market research inquiries to be asked their income or for it to be assessed. Thus, for towns surveyed by market research agencies some

information is available about the incomes of families living in the area concerned. But market research agencies do not cover a sufficient number of towns or people for these data to be available for the country even if they were to be collected together and were otherwise comparable. The only direct information available is thus about the incomes of persons living in a few towns. It accordingly seemed to be worth while trying to find a means of estimating such data from statistics available for other factors that are measured.

Mr. William Coglan of Sales Research Services made available to the writer some results of a sample survey made of about 20,000 households in about thirty towns, which were interviewed in connection with his Brand Barometer[2] service in 1950. His interviews were with the housewives of households selected from the electoral rolls by a method of systematic sampling. In the course of the interview the basic wage rate of the senior wage-earner of the household was obtained, and where this was not known or the information was refused it was possible to obtain a reasonable estimate from the trade union rates or other national wage agreements. These data afforded a good basis from which to start the attempt to find a way of estimating the social class of the larger towns of the whole country.

The first step was to pool the wage data so that the sampling error might be reduced to a minimum, and then to classify the towns for which the sample data were available according to the percentage of

TABLE I

Actual and estimated percentages of households in sample towns in which the chief earner's basic weekly wage rate equals or exceeds £7 10s.

Town	Actual	Estimated	Difference
Brighton and Hove	31	31	0
Oxford	28	31	3
Reading	25	23	— 2
Ipswich	24	23	— 1
Southampton	24	22	— 2
Exeter	23	23	0
Bristol	22	23	1
Portsmouth	22	21	— 1
Leeds	21	20	— 1
Newcastle	21	22	1
Leicester	20	18	— 2
Nottingham	20	20	0
Cardiff	20	18	— 2
Norwich	19	19	0
Birmingham	19	18	— 1
Plymouth	18	18	0
Derby	17	16	— 1
Halifax	17	18	1
Manchester	16	19	3
Sheffield	15	15	0
Liverpool	14	14	0
Coventry	13	16	3
Middlesbrough	9	10	1
			Average difference 1·1%

households where the senior wage-earner had a basic wage rate of £7 10s. a week or more. The businessman might feel inclined to think of this measure as suggesting the proportion of households which had sufficient income to purchase something other than the basic necessities of subsistence. The data are in the first column of figures of Table I.

From this table it will be seen that the towns were scattered throughout the country. The information in the table might be regarded as affording a good indication of the 'social class' of the towns included, but to obtain by direct methods similar information for the 180 or so other large towns would be quite impossible. An equation was, however, evolved which would estimate this measure for the 23 towns of Table I with an average error of about 1 per cent., and all the information needed for this equation is regularly published. The estimated values are also in Table I. An equation which fits the results for these towns so well should provide a very useful estimate for other large towns. Various forms of equation were tried before arriving at the one which provided the results shown in Table I, and these will now be described.

Estimating Equations

The fullest equation was

$$y = 0{\cdot}47x_1 + 0{\cdot}66x_2 + 0{\cdot}20x_3 + 0{\cdot}11x_4$$

where y is the percentage of households in which the senior wage-earner received a weekly basic wage in excess of or equal to £7 10s., x_1 the reciprocal of the infant mortality rate for the town, x_2 the logarithm of the rateable value of 'other hereditaments' a head, x_3 the illegitimacy rate per 1,000 related live births, and x_4 the logarithm of the number of private motor cars a head. All measures are, of course, in standard terms. This equation gives a multiple correlation coefficient of 0·91.

If only the 'best' three items are used, the multiple correlation coefficient is 0·90, and the equation becomes

$$y = 0{\cdot}53x_1 + 0{\cdot}67x_2 + 0{\cdot}21x_3.$$

This equation was used to provide the estimates given in Table I for the sample towns. Table II shows the results of applying the same equation to the 180 larger towns of England, Scotland and Wales.

The 'best' two items give a multiple correlation coefficient of 0·88, and the equation becomes

$$y = 0{\cdot}56x_1 + 0{\cdot}70x_2.$$

Factors studied in developing the Equations

It may be of interest if something is said about the factors used in the equations and a few notes given on some of those which were examined and rejected. Some of the rejected items might suggest alternative forms of estimating equations which could be used for the types of analyses for which the equations given are unsuitable.

TABLE II

Estimated percentages of households in the large towns of Great Britain in which the chief wage-earner's basic weekly wage rate equals or exceeds £7 10s.

Accrington 15	Hornchurch 19	Southampton 22
Ashton-under-Lyne .. 14	Hove 37	Southend 28
Barnsley 13	Huddersfield 17	Southport 25
Barrow 12	Ipswich 23	Stanley 10
Bath 24	Keighley 17	Stockport 15
Bebington 19	Kingston-upon-Hull .. 14	Stockton 11
Bedford 28	Lancaster 17	Stoke-on-Trent .. 8
Beeston and Stapleford 14	Leeds 20	Stretford 24
Birkenhead 14	Leicester 18	Sunderland 9
Birmingham 18	Leigh 8	Sutton Coldfield .. 30
Blackburn 15	Lincoln 22	Swindon 25
Blackpool 22	Liverpool 14	Thurrock 15
Bolton 17	Luton 21	Torquay 31
Bootle 9	Maidstone 24	Tynemouth 12
Bournemouth .. 38	Manchester 19	Uxbridge 25
Bradford 14	Mansfield 13	Wakefield 18
Brierley Hill 9	Middlesbrough .. 10	Wallasey 21
Brighton 26	Newcastle-under-Lyme 10	Wallsend 7
Bristol 23	Newcastle-on-Tyne .. 22	Warrington 8
Burnley 14	Northampton .. 21	Walsall 10
Burton-on-Trent .. 15	Norwich 19	Watford 31
Bury 15	Nottingham 20	West Bromwich .. 11
Cambridge 31	Nuneaton 10	West Hartlepool .. 9
Carlisle 13	Oldbury 12	Widnes 6
Cheltenham 24	Oldham 11	Wigan 8
Chester 19	Orpington 26	Woking 32
Chesterfield 17	Oxford 31	Wolverhampton .. 14
Colchester 24	Peterborough 20	Worcester 18
Coventry 16	Plymouth 18	Worthing 33
Crewe 11	Poole 29	Yarmouth 21
Crosby 19	Portsmouth 21	York 21
Darlington 18	Preston 13	
Derby 16	Reading 23	Cardiff 19
Dewsbury 14	Rochdale 14	Newport, Mon. .. 13
Doncaster 20	Romford 24	Rhondda 6
Dudley 10	Rotherham 10	Swansea 16
Eastbourne 33	Rowley Regis .. 10	
Exeter 23	Rugby 21	Aberdeen 20
Gateshead 9	St. Helens 5	Coatbridge 12
Gillingham, Chatham and Rochester .. 14	Salford 13	Dundee 20
Gloucester 18	Scarborough 30	Dunfermline .. 21
Gosport 22	Scunthorpe 17	Edinburgh 26
Grimsby 18	Sheffield 15	Glasgow 18
Guildford 35	Shrewsbury 27	Greenock 11
Halifax 18	Slough 23	Kirkcaldy 13
Harrogate 27	Smethwick 12	Motherwell and Wishaw 11
Hastings 29	Solihull 25	Paisley 12
	South Shields .. 8	

NOTE. The values for seaside towns are not directly obtained from the equation, since the excess of hotels and boarding houses upsets the rateable value data. The estimates for seaside towns are 'spliced' values. A modified equation was used also for Scotland.

The factors used by Mr. Colin Clark and others in similar exercises before the 1939 war have little meaning today. The rates of unemployment, the proportions of populations who are in receipt of poor relief, and the like, do not now assist very much in determining the characteristics of an area. Unemployment rates and National Assistance data were studied, but rejected at an early stage, since the returns do not cover local authority areas. Unemployment rates are also difficult to

interpret for the reason that local office areas for the Ministry of Labour and those for the National Assistance Board do not coincide, and neither coincides with any other administrative area. Some adjustments were made and tests carried out, but the data were clearly seen to be not worth the effort of handling.

In addition to these factors, the following were studied in detail:

(i) The proportion of persons voting for the Conservative party at the 1950 election.
(ii) The proportion of households in possession of a private telephone.
(iii) The proportion of households owning a private car.
(iv) The illegitimate birth-rate.
(v) Various sickness and mortality rates.
(vi) Various data of rateable value.

The majority of these factors were found to correlate with the chosen criterion of social class: the proportion of households where the senior wage-earner had a basic wage rate of £7 10s. a week or more. The factors which had rateable value as a base had the highest correlation coefficient, and the lowest first order coefficient was with the illegitimacy rate ($r = 0{\cdot}36$). The following are a few notes about the factors taken in turn, those which were examined but which were not used in the final equation being dealt with first.

(*i*) *The proportion of persons voting Conservative* (1950). The first order correlation coefficient between the proportion of votes in the town cast for the Conservative candidate or candidates and the criterion was high at 0·63. Electoral districts do not coincide with those of local authority areas for which population and other important statistics are given, but it was possible to adjust voting figures to local authority areas for the large towns. This was a time-consuming task which proved to be unrewarding, since the inclusion of this factor in the regression equation resulted in a gain of only 0·005 in the square of the multiple correlation coefficient—an insignificant change. It was necessary also to take the logarithm of the proportion voting Conservative in order to transform the data into an unskewed distribution approaching normality.

(*ii*) *The proportion of households having private telephones.* The logarithm of the proportion of households having a private telephone had a correlation coefficient with the criterion of 0·55. But here again the improvement in the equation was not worth the work of obtaining and transforming the data. This may to some extent be due to the effect of business telephones.

(*iii*) *The proportion of registered private cars to people in the town.* It was necessary to take the logarithm of the number of car registrations a head. This factor had a correlation coefficient of 0·46 with the criterion, but in the multiple equation it added almost nothing to the precision of the prediction. This may to some extent be due to the fact that houses having garages are rated higher than those without, and most cars are doubtless garaged on the owner's premises. This factor is

included in the first equation, mainly because it makes some improvement in estimation at the extreme values of the criterion.

(*iv*) *The illegitimacy rate.* It is difficult to explain the first order correlation coefficient of 0·36 between the criterion and the (untransformed) illegitimacy rate for the towns. However, the correlation coefficient was statistically significant, and the inclusion of this factor in the final equation made some improvement. The exclusion of this factor reduced the square of the multiple correlation coefficient from 0·81 to 0·77. The figures given in Table II include this factor, but if the material were being reworked it would not perhaps be worth while to include it.

(*v*) *The infant mortality rate.* The association between infant mortality rates and social class has long been known to medical statisticians. In order to keep all the terms positive and to give an approximately normal distribution the reciprocal of the infant mortality rate was used for expressing this factor. A warning should perhaps be given regarding the source of the basic data. The Registrar-General for England and Wales publishes infant mortality data in two forms: (*a*) the quarterly rate uncorrected for place of birth, and (*b*) the annual rate corrected for place of birth and residence. The uncorrected figures would be meaningless, since areas which have children's hospitals would show a much higher infant mortality rate than is appropriate to residents in that area. This factor had the second highest correlation coefficient of the battery, and, as has been shown, it would be possible to obtain a very good estimate of the social class of towns with this factor and the rateable value alone.

(*vi*) *Rateable value data.* Various ways can be found of building indices from the available data of rateable values. Mr. Gray and Mr. Corlett have used[3] the ratio of 'Industrial' to 'Other' rateable value for towns and rural areas, and have shown that this ratio is very useful as a means of stratifying for sampling such localities. In the present case, however, the best value seemed to be derived from rating data by taking the logarithm of the 'Other hereditaments value' a head of the population of the area. The logarithmic transformation almost completely normalised the data. It will be noted from the third equation that this factor taken together with infant mortality provides a very good estimate of social class, and for most purposes it may not be necessary to include other factors.

Sources of the Data

The rateable value data may be obtained for all the large towns from *Rates and Rateable Values*, published annually by H.M. Stationery Office. Infant mortality data and the information necessary for the calculation of illegitimacy rates are available from the annual reports of the Registrar-General (Medical Section for England and Wales). The number of private car registrations may be obtained only for county boroughs and are available on application to the local authority.

The number of telephones may be obtained from the local telephone directories.

The Uses and Limitations of the Equations

The majority of the limitations of the equations will be self-evident from the commentary on the methods of their construction. For example, all three equations given include infant mortality, and it might therefore be undesirable to use the index of social class for correlating with medical data of other kinds. The medical factor under study might be causally associated with the same factors as infant mortality, and thus produce a spurious correlation with social class.

There is, however, a more serious limitation. The equations were derived on 1949–50 data of rateable value, infant mortality, illegitimacy, and car registrations. Shortly*, the new basis for the rating of property will be in force, and any system based on the old methods of rating will cease to hold. The number of car registrations would be considerably affected by petrol rationing should this again become necessary. Nothing can be said about how long the equations might hold good, and it is suggested that the most useful function they can fulfil is as an estimate pending the publication of the 1951 census data. In this respect they may prove of value to businessmen and economists. To the former they may be used for correlating the social class estimates with sales data or for selecting areas for the pre-testing of products or for similar purposes. To economists they may have value in providing estimates of regional incomes or in the analysis of certain forms of family budget data and the like.

It should be noted that the index relates to the resident population of the local authority area named. Areas such as seaside towns where there is an influx of purchasing power from neighbouring areas require special consideration when economic or marketing research is considered in relation with the index.

* Since this article was written it has been stated in the House of Commons that the adoption of the new basis of rating may be delayed.

REFERENCES

[1] *The Marketing Survey of the United Kingdom* (1950). Business Publications Ltd., London.

[2] *The Brand Barometer*. Sales Research Services. Quarterly (subscribers only).

[3] Gray, P. G., and Corlett, T. (1950). 'Sampling for the Social Survey,' *J. Roy. Statist. Soc.*, A, **113**, 150.

A USE FOR THE JURY QUALIFICATION IN SAMPLE DESIGN

THOMAS CORLETT, M.A.
The Social Survey

The proportion of the electors liable for jury service is an index of economic status of an area that can be determined from the Register of Electors for any area, small or large, definable in terms of streets or roads. It is useful in designing sample schemes.

THE great majority of social or marketing surveys in this country, if conducted on a national scale, involve at some stage the selection of a sample of areas within which samples of people or households are selected. And since areas of any type are known to differ widely in the economic status of their inhabitants—probably the most important single factor affecting the results of a wide variety of such surveys—there naturally arises a demand for indices by which the economic status of areas may be estimated in advance, in order to ensure that the correct proportions of areas of high and low economic status are included in the sample.

But the use of any of those indices hitherto published places a somewhat troublesome limitation on the population sampler. Their values can at best only be calculated for complete local government administrative districts—boroughs, urban districts and rural districts—so that in order to take advantage of them one is almost bound to use these administrative districts as sampling units at the next-to-last stage and to select the sample of individuals or households throughout the whole of each selected district. Now many administrative districts cover large areas—for example, 32 per cent. of the population of England and Wales live in administrative districts over 30 square miles in area. Thus unless a national sample is concentrated in relatively few areas the interviewers' calls on a random sample in an administrative district are often widely scattered, with resulting high expenditure of time, money and morale.

This difficulty would be largely overcome if one had an index of economic status which could be readily calculated for the smaller areas, such as wards and parishes, into which administrative districts are divided. The absence of such an index has in the past led some workers to venture upon the treacherous paths of purposive selection, perhaps beginning with a stratified random sample of administrative areas but within them selecting 'working class' or 'middle class' wards by subjective judgment. The purely random selection of wards, on the other hand, though theoretically preferable, contributes appreciably to the

sampling error, since the wards of most administrative districts vary considerably in economic status.

As the author has shown elsewhere,[1] a solution to this difficulty is available in the shape of an index which can be calculated for any ward or parish, or indeed for any definable area in England and Wales, whether in administrative use or not. The suggested index is a very simple one: it consists of the proportion of the registered electors in any area who are liable for jury service. The names of people liable for jury service are marked in the Register of Electors with the suffix —J. The value of the index for any area can thus be calculated by counting the —J suffixes in the part of the Register covering the area and expressing this number as a percentage of the total electorate of the area. The main advantage claimed for this J-Index, as it has been called, is simply this fact that it is not, like most other indices, limited to any particular kind of area, since its value can be got for any area which is definable in terms of streets or roads.

But why should the percentage of jurors in an area serve as an index of economic status? The chief reason lies in the fact that the most important qualification for jury service is to be a ratepayer living in property of not less than a certain rateable value, the lower limit being £30 in the counties of London and Middlesex and £20 in the rest of England and Wales.

The existence of a correlation between rateable value and income is well established, and additional evidence (given in more detail in a Social Survey report[1]) confirms the usefulness of the J-Index. For example, a random sample of people marked —J in the current Register of Electors was found to contain about three times as large a proportion of people in the upper income group as a sample of the general population, the 'upper income group' being defined for this purpose as members of households the head of which had a net weekly income of £10 or more. Again, in a sample of 22 administrative districts of England and Wales outside the counties of London and Middlesex the observed coefficient of correlation between sample estimates of the proportions of people in the 'upper income group' (similarly defined) in these districts and the values of the J-Index was + 0·62. This figure almost certainly underestimates the true value of the coefficient, since the estimates of the proportions in the 'upper income group' were based on rather small samples. A similar test for ten administrative districts in London and Middlesex gave a coefficient of + 0·83.

The application of this index makes possible the use of a two-stage national sample design for England and Wales with wards and parishes, efficiently stratified for income, as first-stage units. Besides avoiding the disadvantages of using the often inconveniently large administrative districts as sampling units, such a design would have the advantage of greater flexibility in the geographical stratification it would allow. One could, for example, treat separately each of the six officially defined

conurbations, representing each by a stratified sample of wards. When using complete administrative districts as sampling units this is virtually impossible except in the case of Greater London, since each of the other conurbations consists of one or more very large administrative districts together with a number of much smaller ones.

But although the primary purpose for which the index was evolved is its application to sample design, it seems likely that its distinctive property of being adaptable to any kind of area will make it of use for other purposes as well. It might, for example, provide an index of economic demand for areas that in some connections are more significant than local government districts.

REFERENCE

[1] Gray, P. G., Corlett, T., and Jones, Pamela (1951). *The Proportion of Jurors as an Index of the Economic Status of a District.* London: The Social Survey.

A STATISTICAL APPROACH TO THE SPECIFICATION OF PLASTICS

CHARLES WAINWRIGHT, B.Sc., F.R.I.C.

Although this article is primarily concerned with plastics, the underlying general principles apply to other materials. The usual type of specification, in which each batch of material is accepted or rejected according as its tested quality exceeds or falls below a specified level, presents difficulties. These are overcome in the new B.S. 771 for moulding material by regarding groups of batches as wholes and taking account of variability in quality as well as average levels. Mr. Wainwright proposes a somewhat similar but simplified form of specification suitable for laminated sheets.

THE use of specifications prescribing the levels of quality of industrial products is of growing importance. These levels of quality may be laid down in a British Standard, in a purchasing specification drawn up by a user, or in a manufacturing specification used by the manufacturer of the product. With one or two notable exceptions such specifications attempt to assess the quality of a batch of material by the application of standardised methods of test to a sample or samples taken from the batch, and the comparison of the results of the tests with corresponding qualifying limits laid down in the specification.

Now although this is the conventional procedure in general application throughout the world for a multiplicity of goods, in its application to a product of variable quality it can be fully effective only if (*a*) every batch or consignment is tested and (*b*) the quality of the batch can be determined substantially without error.

Limitations of the Conventional Procedure

In the manufacture of many variable products, which may be of high quality, requirement (*a*) cannot be satisfied, as it is frequently neither economically desirable nor physically possible to test every batch. It may be excessively expensive to examine any more than one batch out of each hundred manufactured.

Requirement (*b*) is often difficult to satisfy, as each batch is usually appraised on a sample, thereby adding sampling errors to such testing errors as may exist. In many instances, for a practicable size of sample, the errors can nullify any attempt to assess the quality of an isolated batch and thus make it impossible to decide, with a reasonable degree of certainty, whether the batch meets the requirements of a specification, unless it does so with a very large margin.

Because of the testing and sampling errors, even with a sequence of

batches of precisely the same level of quality $\bar{X}$, the test results would be as shown in Fig. 1. If the user required material of true quality not worse than the level represented by L and, ignoring the sampling and testing errors, insisted upon a specification limit equal to L, the manufacturer would have to produce a material of true quality equal to $\bar{X}$ because he could not run the risk of arbitrary rejection. The user would therefore receive material of higher quality than he required and would have to pay more for it. Moreover in advertising his product

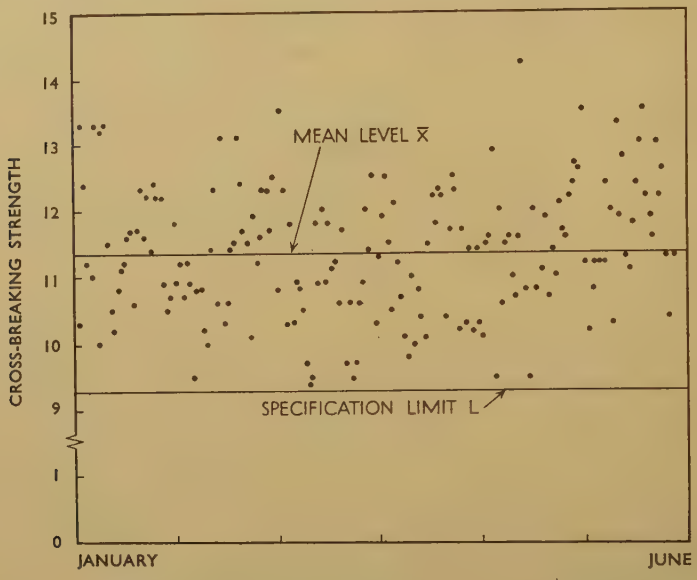


FIG. 1. Cross-breaking strength (1,000 lb./sq. in.) of a sequence of batches of a moulding material of uniform quality.

the manufacturer would not be able to do justice to its quality if he could claim only that the worst result likely to be obtained on it would not fall below the sufficiently safe, and therefore relatively low, specification level.

A further and serious drawback to the conventional single-limit specification is that it does not permit of the adequate comparison of materials. Materials purchased in accordance with the requirements of such a specification may differ very significantly in average quality, as shown in Fig. 2.

These figures are the results of the measurement of insulation resistance on many batches of two types, A and B, of synthetic resin bonded laminated sheet materials. Although both materials may be claimed to satisfy the requirements of the specification, the average quality of material A is very much better than that of material B, and on this account material A is likely to give a much better performance in service. It is sometimes argued that the greater variability of material A outweighs any advantages which may accrue from the fact that the average quality is high, but for many materials this is not so.

The weakness of the conventional type of quality specification as described above does not appear to be sufficiently well understood by those dealing with the manufacture and the purchase of materials of variable quality, or materials for which the variation inherent in the method of test is appreciable. A considerable amount of controversy can and does arise from this lack of understanding.

The levels of quality in specifications of the single-limit type are usually arrived at by agreement between manufacturer and user, but

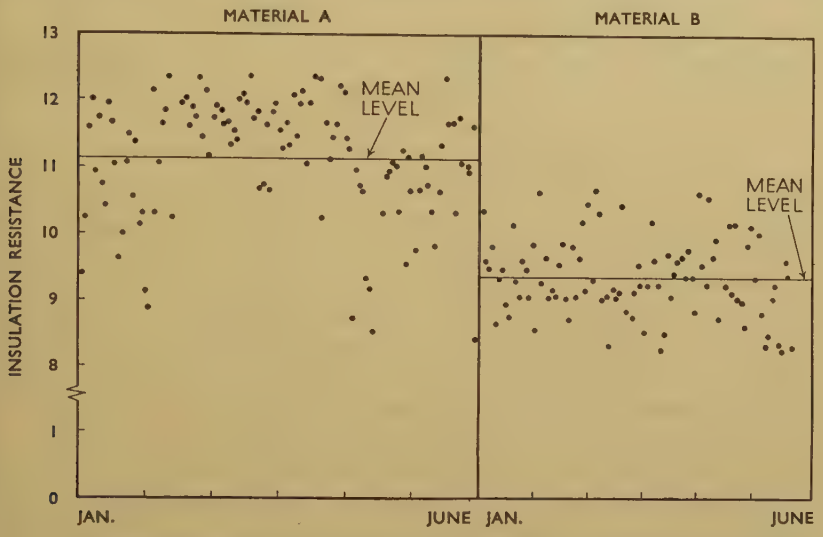


FIG. 2. Insulation resistance ($\log_{10}$ ohms) of a sequence of batches of two grades of synthetic resin paper-filled laminated sheet.

very often neither party is completely satisfied with the outcome because of a lack of appreciation of the problems involved. The user may occasionally test a few batches of a product which he has found over a period of years to be satisfactory for his purpose. He decides to write a purchasing specification based upon tests on say ten deliveries, which would perhaps just make all of them acceptable; or he may make use of the information at his disposal to assist, with the manufacturer, in the preparation of a British Standard. The manufacturer is invited to comment on a draft of the specification and finds that the levels proposed by the user would reject say 5, 10 or even 20 per cent. of the hundreds of batches of the same grade of material tested by him.

Although the manufacturer may have tested several hundred batches of material in the course of a year, they may nevertheless represent not more than a few per cent. of all the batches made during that year. He therefore hesitates to give any guarantee for the grade as a whole, and hence for single batches or deliveries, unless the user will amend his specification and agree upon a single qualifying limit set at a level of quality which will reject not more than about 1 per cent. of the material manufactured. This is a reasonable attitude, but the

purchaser can be excused for assuming that *his* attitude, backed by the information at his disposal, is equally reasonable. What is the way out of this apparent impasse, and how can a material be properly described and specified, for which the batch quality can only be determined with an error and which, moreover, may vary from batch to batch ?

The Statistical Approach to Quality Specification

It is possible to find a solution if the problem is reduced to its simplest terms. A grade of material as a whole consisting of many hundreds or even thousands of batches may have been found during a given period of time to be wholly or almost entirely acceptable to the users of the material. The quality of the grade during the given period can be shown as a series of numerical values, each representing a separate batch and scattered around an average value. (As each value is preferably the mean of several determinations it can conveniently be referred to as the 'batch mean.') Alternatively, these numerical values may be shown graphically in a simple dot diagram similar to Fig. 1. The scatter of the batch means can usually be attributed to the operation of two factors corresponding to between-batch and within-batch variability. Between-batch variability arises from the true variation in quality of the material from batch to batch. Within-batch variability is variation between different test results obtained for the same batch, and arises from variation due to testing errors and from real variations in quality from place to place within the batch. Within-batch variability is the source of the error with which the mean quality for each batch is determined, and the extent of this error depends upon the extent of within-batch variability and upon the number of tests per batch.

The plastics industry has already sponsored a British Standard (B.S. 771 : 1948, 'Synthetic Resin (Phenolic) Moulding Materials') in which the average quality of a sequence of batches and the within-batch variability are specified. In this Standard it is virtually recognised that effective inspection of every batch cannot be economically attained and that the quality of the material can properly be assessed only by a survey of the results obtained over a period of time on many batches. This British Standard represents a complete departure from the conventional type of single-limit specification in that it lays down standards of quality appropriate to the results obtained on a series of batches over a period of time, while specifying qualifying limits below which the results obtained on not more than 1 per cent. of batches would be expected to fall. The new B.S. 771 has proved to be a very considerable improvement on the Standard it replaced, which was of the conventional single-limit type, and its acceptance by both manufacturers and users of the product has justified the use of its particular statistical technique.

B.S. 771 lays down standards for the mean and standard deviation,

but it should be particularly noted that the latter requirement covers only within-batch variability and not between-batch or total apparent variability. The justification for this is that in this instance the within-batch variability (substantially the test variability) often accounts for a predominating proportion of the total apparent variability between batches. It would be an improvement, on theoretical grounds, to include an additional requirement in the specification which would restrict the allowable amount of true between-batch variability, but this could easily give rise to a totally unprofitable complexity. The technique adopted in B.S. 771 therefore represents a compromise with some concessions to expediency. It is nevertheless a reasonably satisfactory compromise, suited to the nature of the material to which it is applied.

The method of B.S. 771 as described above is not altogether suited to certain other products of the plastics industry. An example is laminated sheet made from paper or fabric impregnated with synthetic resin. The product in question consists of material in sheet form, and a proposed British Standard to cover it will lay down standards of quality for five or six different physical properties. In practice a relatively large amount of testing, mostly destructive, is carried out during the course of manufacture in order to check the quality of the grade as a whole. Nevertheless not more than a small fraction of the total number of sheets manufactured in a given period is subjected to test. Any specification acceptable to the manufacturer must deal with the grade as a whole, while preferably making some provision for dealing with individual batches. It should be noted here that any definition of batch must be entirely arbitrary, and here we regard each sheet as a batch. Each type of material is made in a range of thicknesses, and in some cases the results of test are influenced by the thickness. Thus the large variety of sheets involved and the large number of properties which have to be measured to assess the quality of a given sheet necessarily tend to make even the simplest possible type of specification rather unwieldy. To avoid adding further complication and thereby devising a specification unacceptable to industry it is essential to keep any statistical technique as simple as is compatible with effectiveness.

Simplification of the Specification

With this material simplicity may be achieved by taking advantage of the fact that there is no special need to distinguish within-batch from between-batch variability. In testing this material it is not physically possible to isolate, as in B.S. 771, the test variability because the total apparent variation within a single sheet may be due in part to the effect of real variation from place to place and in part to test variability. It would be possible, nevertheless, to specify within-sheet variability or between-sheet variability separately or both together. The last proposal is again unacceptable on account of the complexity of the resulting specification, having in mind the variety of types of

sheet and the relatively large number of properties to be specified, and realising that provision has to be made to accommodate the effect of thickness on test results. It has been found, moreover, with the material under discussion that the between-sheet variability is often greater than the within-sheet variability, and in this respect the material differs from that covered by B.S. 771. In use it is almost always cut into small pieces, so that to the user the variation within a sheet is of no more importance than the variation between sheets.

Having all these points in mind it becomes apparent that a sufficiently complete picture of quality is given by the dot diagram. Then, provided that the number of tests per batch is specified, the quality of the material as a whole can be described adequately by a mean level of quality representing the mean of all the results and by some measure of the amount of scatter of the batch means about this grand mean. The most suitable measure of scatter in this instance is the standard deviation of the batch means, and this standard deviation together with the mean level of quality would form a satisfactory specification. As the standard deviation in this instance covers the total scatter of batch means it is not concerned with the separate factors responsible for the scatter.

It is possible to make yet one further concession for the sake of practicability. A specification of the type just described would give a level of $\bar{\bar{X}}$ for the average quality of a large series of batches and a level σ for the standard deviation of a large series of batch means about $\bar{\bar{X}}$. In accordance with statistical theory, the required minimum average level for any given number of batch means would be

$$\bar{\bar{X}} - \frac{k\sigma}{\sqrt{n}}$$

where n is the number of batches and k depends upon the acceptable risk of chance alone giving a batch mean below this value. (For a risk corresponding to a probability level of 0·01, $k = 2{\cdot}33$.)

Similarly, the required standard deviation for any given number of batches can be calculated by conventional methods. It is suggested, however, that the specification and assessment of the quality of the material under review would be sufficiently effective if the direct test for the standard deviation of a series of batch means were omitted. The justification for omitting this test is that the allowable level of average quality appropriate to a given number of batch means is related directly to the value of standard deviation laid down in the specification, as shown by the above formula. While this relation is strictly true for any number of batches its practical value as a check on standard deviation is limited to cases where n is small. For example, the required level for one batch is $\bar{\bar{X}} - k\sigma$ and is thus very much dependent upon the value of σ, while if n is very large the level becomes very close to $\bar{\bar{X}}$ and is virtually independent of σ. As the level for a single batch is highly dependent upon the specified standard deviation, this level can

be used for determining whether the standard deviation of a series of single batch means is greater than the specified standard deviation. If, therefore, the specification contains a requirement that the level of quality of single batches shall not fall below $\bar{X} - k\sigma$, any material of average level $\bar{X}$ which had a greater standard deviation of batch means than the specified value would in due course fail to comply with the specification.

A possible draft of a specification drawn up in this manner is given below.

> A grade of material shall be deemed to comply (or not to comply) with this Standard if the mean of the results of tests on a single sheet, and on a sequence of any number of sheets, is not worse (or is worse) than the appropriate qualifying limit calculated from the basic values for mean and standard deviation in accordance with the expression:
>
> $$\text{Qualifying limit} = \bar{X} - \frac{2{\cdot}33\sigma}{\sqrt{n}}$$
>
> where $\bar{X}$ is the basic value for mean,
>
> σ is the basic value for standard deviation,
>
> and n is the number of sheets.

It has been pointed out that a specification of this type would reject a grade of material of average quality equal to the specified value but with standard deviation higher than the specified value. It is clear that it would reject a grade with low average quality, as, even if the standard deviation of such a grade were so low that single batches did not tend to fall below the specified value, the average level of a sequence of batches would soon fail to meet the appropriate requirements. There is finally a type of material, in which both the average level of quality and the standard deviation are higher than the specified values, which must be given special consideration, particularly as it represents a significant proportion of synthetic resin laminated sheet of acceptable quality. Without going into detail it can be shown that the specification outlined above would not reject such material unless the standard deviation were higher than the specified value by an amount exceeding $q/2{\cdot}33$, where q is the amount by which the average quality is better than the specified value. This implies that for material of higher average quality than $\bar{X}$ the extent of scatter of batch means can be greater than that corresponding to σ, provided that no single results fall below the minimum level set for single batches of material corresponding exactly to the specification requirements for $\bar{X}$ and σ. The manufacture and use of materials of this type are well established, and in some instances their high average level becomes of special value to the user in spite of the higher standard deviation. From the practical point of view, therefore, it is an advantage rather than a disadvantage that the suggested specification will admit such material as satisfactory.

Certification to a British Standard

When a British Standard is used as the basis of a formal scheme of guarantee of quality (a certification scheme) under the auspices of the British Standards Institution it becomes particularly important that there should be a precise definition of conformity with the Standard acceptable to both user and manufacturer. In cases where 100 per cent. inspection has been carried out, i.e. where every certified article has been tested, the implications of certification are fairly clear. Where a proportion only of the certified goods has been tested, and where the grade as a whole is certified, it is essential that the terms under which permission to certify is granted shall bear close scrutiny and shall be based upon a definition of conformity laid down in the appropriate British Standard.

A single-limit specification would not adequately describe the properties of laminated sheet and could not properly serve as the basis of a certification scheme unless certification were confined to those sheets which had been tested or unless the specification were supplemented with a statistical sampling scheme. It is difficult with this particular material to devise a statistical sampling scheme which would be suited to the various manufacturing techniques in current use and which could at the same time be applied to a series of consignments received by a user. Another difficulty is that the variability within a sheet will introduce a source of uncertainty in appraising the quality of a sample and hence of a batch. This source of error can reduce considerably the value of assessing the quality of a batch.

It is suggested that the type of specification advocated here would be eminently suitable as the basis of a certification scheme because:

(*a*) It describes the material as a whole and can therefore be used effectively where only a proportion of batches is tested.

(*b*) It does not need to be supplemented with a statistical sampling scheme.

(*c*) It is worded in such a way that it is possible to obtain evidence of conformity or non-conformity not only from the small number of tests which may be carried out by a purchaser on consignments received by him but also from the very much larger number of tests normally carried out by the manufacturer in the exercise of technical control during the course of production. This feature is of considerable value to the manufacturer as it avoids duplication of testing, and delay and inconvenience arising from the sampling of material which has been assembled into consignments and awaits despatch.

(*d*) The form of the specification lends itself to the use of statistical quality control charts, from which evidence of conformity or non-conformity is readily apparent.

USES AND ABUSES OF FACTOR ANALYSIS

HANS J. EYSENCK
Psychology Department, Institute of Psychiatry, Maudsley Hospital

Factor analysis, if used cautiously and with a full understanding of its underlying logic, can illuminate many fields of research. It is a statistical method, based on correlation analysis, for combining a number of measured variables into a smaller number of groups of variables: the groups are the factors. They are in no way absolute and the basis of their choice is partly subjective: they constitute a hypothesis that requires to be tested. Examples are given to show the fruitfulness of factor analysis.

Professor T. L. Kelley, in his *Fundamentals of Statistics*,[6] has pointed out that 'there are two occasions for resort to statistical procedures, the one dominated by a desire to prove a hypothesis, and the other by a desire to invent one. This has led to distinct schools of statisticians, both lying within the general field of scientific endeavour.' Factor analysis may be used for either of these purposes; it may be useful in suggesting hypotheses, and it may be useful in testing hypotheses.

If factor analysis is not differentiated from other statistical methods by its purpose, we must look for some other distinguishing mark. This has been provided by Professor M. G. Kendall's[7] division of methods of multivariate analysis into *analysis of dependence* (analysis of variance and covariance, regression and confluence analysis, discriminant analysis), and *analysis of interdependence* (correlation analysis, association and contingency analysis, component analysis). 'In the latter we are interested in how a group of variates are related among themselves, no one being marked out by the conditions of the problem as of greater prior importance than the others, whereas in the analysis of dependence we are interested in how a certain specified group (the dependent variates) depend on the others.' Factor analysis plays its part in the analysis of interdependence, and is indeed a logical and historical outgrowth of the work of Galton and Karl Pearson on correlation analysis. (It is not always realised that we owe the original formulation of factor analysis to Pearson and not, as is often believed, to Spearman.)

Statistical procedures are created to solve certain scientific problems; the nature of these problems must be clearly understood if we are to avoid misapplication of such procedures. Most orthodox statistical methods give an answer to the question of the level of confidence at which a given set of differences disproves the null hypothesis; this type of question arises naturally in fields where contrasted groups and

treatments can be clearly and easily identified. Such methods are widely used in psychology when questions are raised involving sex or class differences, the effects of certain types of therapy or education, or the prediction of success or failure on clearly denoted criteria. Unfortunately, not all questions which may be legitimately raised can be answered by means of these orthodox procedures, and indeed some of the most fundamental questions in psychology precede in point of time the possibility of applying the procedures. In order to apply the *t*, *F*, or *chi squared* tests or any of their equivalents we must have identified the groups (whether of people or of test scores) which are to be compared or contrasted. When the knowledge implied in the setting up of such groups is not available, we face problems which non-factorial methods are not designed to handle, and statisticians whose hostility to factor analysis blinds them to the nature of the problem which this technique sets out to solve do a disservice to science by pretending that orthodox procedures can give answers to questions they were not designed to deal with.

Fundamentally, it may be said that factor analysis attempts to reduce the multiplicity of observed uniformities of nature to a much smaller number of hypothetical causes or factors. As Professor Holzinger and Mr. Harman[5] put it: 'Factor analysis is a branch of statistical theory concerned with the resolution of a set of descriptive variables in terms of a small number of categories or factors. This resolution is accomplished by the analysis of the intercorrelations of the variables. A satisfactory solution will yield factors which convey all the essential information of the original set of variables. The chief aim is thus to obtain scientific parsimony or economy of description.'

These factors are of course abstract and may justly be called 'statistical artefacts'; they are artefacts in the same way that any scientific concept (electron, aether, energy) is an artefact. Such factors may or may not have a physiological basis; the demonstration, for instance, that the factor of 'neuroticism' or 'emotional instability' is very largely inherited, and inherited as a whole, links this product of factor analysis with important facts and theories in other sciences, and thus establishes it more securely.[3] But in order to prove the usefulness of a factor such a demonstration is not necessary and may be impossible; we are unlikely to discover physiological factors that relate to patterns of interests or attitudes. Nevertheless such patterns can be shown to exist, and we must find some method of 'parsimonious description' which will allow us to handle these data scientifically.

In a sense, what factor analysis may be said to attempt is the dimensional analysis of its particular field of study; it seeks to isolate, identify, and measure objectively the main dimensions along which the units of study may be located. It is often objected that this procedure implies a certain subjectivity of choice; the results achieved by one analyst may differ from those achieved by another analysing the same table of intercorrelations.

In actual fact results reported by competent workers are usually remarkably congruent; thus, in the intellectual field, very much the same factors have been isolated by Professor Thurstone,[8] Professor Holzinger,[4] and Sir Cyril Burt,[1] using quite different methods of analysis. In any case a certain amount of subjectivity is a feature of all dimensional analysis and is not in any way exclusively connected with factor analysis. We may quote in support the physicist Dr. Bridgman, who maintains that 'there is nothing absolute about dimensions . . . they may be anything consistent with a set of definitions which agree with the experimental facts.' If the choice of dimensions is up to a point arbitrary in physics, it seems unreasonable to expect factor analysis or any other statistical procedure to give us psychological dimensions which are not equally arbitrary. All possible systems of dimensions to describe a given set of facts must be convertible into each other, as they must all agree with the experimental facts; if two systems of dimensions disagree an empirical test becomes possible to decide which of the two leads to deductions verifiable by experiment. A clear example of such experimental clarification is presented by Thurstone's disproof of Spearman's contention regarding the sufficiency of one single factor to account for the intercorrelations of cognitive tests. This example may serve as proof—if such be needed—that factor analysis can be used in the two ways suggested at the beginning of this paper. Spearman's factorial studies suggested to him the 'single factor theory' (factor analysis suggesting a hypothesis), and Thurstone showed conclusively that this hypothesis was not tenable (factor analysis disproving a hypothesis). This disproof led to new hypotheses (factor analysis suggesting a hypothesis) which to date have in the main survived experimental tests (factor analysis supporting a hypothesis).

Two examples may show the way in which the field of research may be clarified by means of factor analysis. The first example will be taken from the field of physiology. Dr. Wenger[9] put forward the hypothesis that in some people the sympathetic branch of the autonomic nervous system was dominant, in others the parasympathetic branch, and that in the majority a balance existed between these two. Hypotheses similar to these have been held by many physiologists before him. By taking a number of measures of physiological reactions known to be determined, some by sympathetic and others by parasympathetic innervation, and by correlating these for groups of children, adolescents, and adults, he showed that the factor loadings which emerged from the analysis of the correlations agreed precisely on the tests specified by his hypothesis.

The other example is taken from my own work[2] in the field of social attitudes. For a long time two apparently contradictory hypotheses have been held by many sociologists and social psychologists regarding the structure of social attitudes. On the one hand, Communists and Fascists were considered to lie to the left and right respectively on a continuum towards the centre of which lie the Socialists, Liberals, and

Conservatives in that order. This would put Communists and Fascists at opposite poles. On the other hand, it is often held that Communists and Fascists are very similar in many ways and that there may be little to choose between them. A factor analysis of the correlations between a number of social attitudes revealed two main factors; one of radicalism-conservatism; the other tentatively labelled 'tough-mindedness *versus* tender-mindedness.' It was shown, by giving factor scores to members of the Fascist, Communist, Socialist, Liberal, and Conservative parties, that these parties could not be represented in terms of a single continuum (running shall we say from left to right ?), but that while the Communists and Fascists were respectively very radical and very conservative, they were also extremely tough-minded; Socialists and Conservatives were respectively radical and conservative in their outlook and neither very tough- nor very tender-minded. Liberals were neutral with respect to the radical-conservative dichotomy but tended to be remarkably tender-minded. Thus the factor analysis reconciles the two original hypotheses (Communists and Fascists are at opposite poles with respect to the radicalism factor but are very similar with respect to the tough-mindedness factor) and gives a much better description of the field of social attitudes organisation than was previously available.

These, then, are the uses of factor analysis; what are the abuses to which it has given rise ? In the main these would seem to follow from a failure to make the distinction suggested between the two main functions of factor analysis. Many writers conduct a factorial study the results of which suggest a hypothesis and then proceed as if the hypothesis was thereby proved. This, of course, is completely inadmissible, and much of the criticism which factor analysis has provoked is due to this elementary error. It would not seem reasonable, however, to condemn a scientific method because some of its junior practitioners have misunderstood the logic underlying it. Another abuse which has given rise to criticism is the occasional tendency of some writers to treat the results of factorial studies as being in some mysterious sense absolute and fundamental; this lack of philosophical sophistication fortunately does not characterise any of the well-known factor analysts in their more carefully considered pronouncements, and while quotations can be torn out of their context to give a contrary appearance there can be no doubt that the relative nature of factorial constructs is appreciated by all of them. A third abuse is due to the absence of suitable and usable criteria of significance in factor analysis; in their absence data are often over-interpreted, and chance fluctuations are made the basis of far-reaching conclusions. This is a very real fault, and every effort should be made to find methods for the estimation of errors; it is here that the help of statisticians would be most appreciated. Duplication of experiments provides a reasonable safeguard against fallacies attributable to this difficulty, but again many analysts fail to guard against criticisms on this score.

In summary, it may be said that factor analysis is a complex and difficult method which should be used with great caution, and with a full knowledge of its underlying logic; those who use it as a sausage machine which will grind out absolute truth regardless of the data which are fed into it or of the underlying imperfection of their research design, fully merit the criticisms often levelled against factor analysis as a method. But these criticisms do not apply to those who appreciate fully the many pitfalls which beset any statistical evaluation of complex data, and who have chosen factor analysis after careful consideration as the only method at present available to solve certain urgent problems fundamental to their science.

REFERENCES

[1] BURT, Sir C. (1940). *The Factors of the Mind.* University of London Press, London.

[2] EYSENCK, H. J. (1951). 'Primary Social Attitudes as related to Social Class and Political Party,' *Brit. J. Sociol.*, **2**, 198–209.

[3] EYSENCK, H. J., and PRELL, D. B. (1951). 'The Inheritance of Neuroticism: an Experimental Study,' *J. Ment. Sci.*, **97**, 441–65.

[4] HOLZINGER, K. J. (1934–6). *Preliminary Reports on the Spearman-Holzinger Unitary Trait Study, Nos.* 1–9. Chicago University Press, Chicago.

[5] HOLZINGER, K. J., and HARMAN, H. H. (1941). *Factor Analysis: a Synthesis of Factorial Methods.* Chicago University Press, Chicago.

[6] KELLEY, T. L. (1947). *Fundamentals of Statistics.* Harvard University Press, Cambridge.

[7] KENDALL, M. G. (1950). 'Factor Analysis,' *J. R. Statist. Soc.*, **B**, **12**, 60–73.

[8] THURSTONE, L. L. (1938). *Primary Mental Abilities.* Chicago University Press, Chicago.

[9] WENGER, M. A. (1948). *States of Autonomic Balance in Army Air Forces Personnel.* University of California Press, Berkeley and Los Angeles.

THE ACCURACY OF SYSTEMATIC SAMPLING FROM CONVEYOR BELTS

GEOFFREY H. JOWETT
University of Sheffield

The sampling of coal and similar materials for testing purposes is described, and it is shown how the results of a trial set of observations may be used to calculate the accuracy of a given sampling scheme. The article presents an interesting practical application of the analysis of variance and a new technique based on the theory of time series.

Since it is difficult to take a random or representative sample from such a bulk as a truck of small coal or a pile of iron ore, it is customary to sample such material while it is still on the conveyor belt if this is at all possible. Shovelfuls, technically termed 'increments,' are usually taken from the belt at intervals, and freedom from bias is secured by including in the increment a complete cross-section of the belt. On theoretical grounds there is much to be said for selecting the increments randomly along the belt, or for dividing the stretch of belt corresponding to the truck or pile into equal lengths or 'zones' and choosing an increment at random from each zone (such techniques are forms of area sampling and have been well discussed). But such a method is usually impracticable; it is easy to instruct the sampler to stand at a fixed point and take increments at fixed time intervals, but virtually impossible to instruct him to take them at random or stratified random intervals of time. Accordingly samples from conveyor belts are usually taken at fixed time intervals, so that the increments are effectively evenly spaced along the belt and constant in number for similar trucks. The increments when taken are usually combined into a single sample and the mean ash content of the coal, or whatever variable may be of interest, is estimated from this single sample. This sample is large in bulk and is reduced by various methods until a sample small enough for chemical analysis is obtained. The ash content of this sample may be regarded as the mean of the ash contents of the separate increments. We are not concerned here with the effectiveness of this reduction, which can be investigated by standard statistical methods, but with the sampling error with which the mean value for the increments included in the sample estimates the true mean value over the whole stretch of belt corresponding to the truck. This sampling error cannot be assessed from the sample itself: it must be predicted from some kind of experiment carried out on typical stretches of belt. Such an experiment may take the form of a systematic sample for several typical stretches of belt, with increments taken at much closer intervals than

normal and with separate analysis of each increment, or alternatively of a number of 'interpenetrating' samples (e.g. four systematic samples each a quarter of an interval out of phase with the next) for a number of typical stretches of belt, the increments for each such sample being combined before analysis. If a number of uniformly spaced increments are numbered 1, 2, 3, . . . in order, four interpenetrating samples would be formed by combining respectively increments

1, 5, 9, . . . 2, 6, 10, . . . 3, 7, 11, . . . 4, 8, 12, . . .

The data resulting from an experiment have in the past usually been treated by analysis of variance methods in order to provide a prediction of the standard error of the mean of any systematic sample with a given interval between the increments. With this approach it is necessary to design the experiment to fit in with this given interval, and it is difficult to predict the consequence of taking an interval of different size. The modern theory of time series, however, can be applied not only to give an equally ready solution to the problem when the two intervals' sizes correspond, but also to give a solution when they do not. From a single experiment the standard error of the mean obtained by taking different intervals between the increments (and hence different total numbers of increments per truck) can be estimated and compared.

The time series technique is preferable on theoretical grounds because it is based on more realistic assumptions than the analysis of variance approach, and on practical grounds because it is more useful for designing a sampling scheme and deciding on the right interval to take. It is also of statistical interest as a simple new technique depending on the modern theory of time series; examples of such techniques are still rare. On the other hand, when it is desired only to assess the accuracy of a specified systematic sample, the analysis of variance method may be preferred because of its greater familiarity and independence of graphs.

Analysis of Variance Method

There is considerable variation in the way in which analysis of variance may be applied, but the two methods described below are typical. A set of data for illustrative purposes is given in Table I.

Method A. The data for this method are the ash contents for individual increments, given in the top part of Table I. The conveyor belt is working continuously, loading one truck after another. For four trucks (in this case not consecutive) increments have been taken from the belt at one-minute intervals while each truck was being loaded, giving twenty increments for each truck. For the purpose of the analysis the total length of the delivery to each truck has been divided arbitrarily into zones, each zone consisting of the length delivered during four minutes. The variation within each zone is presumed to be substantially random, and the analysis of variance is performed as in Table II to obtain a pooled estimate of this for all zones. It is

TABLE I

Illustrative data: Ash contents of increments taken at one-minute intervals from conveyor belt sections loaded into four trucks

A. Individual Increments

Zone No.	Increment No.	Truck 1	Truck 2	Truck 3	Truck 4
1	1	4·80	5·11	5·42	5·82
	2	4·96	4·58	5·32	5·91
	3	4·98	4·98	5·26	6·20
	4	4·58	4·43	5·17	6·12
2	5	4·54	4·75	5·19	5·70
	6	4·51	5·20	5·38	5·94
	7	4·29	4·95	5·27	6·03
	8	4·55	5·11	5·54	5·90
3	9	4·45	5·26	5·52	6·06
	10	4·28	5·22	6·26	5·73
	11	4·10	5·07	5·92	5·77
	12	4·30	5·16	5·53	5·68
4	13	3·85	5·22	5·56	5·93
	14	3·83	6·08	5·61	5·81
	15	3·94	5·53	6·07	5·60
	16	4·01	5·56	5·61	5·84
5	17	3·87	5·40	5·52	5·47
	18	3·88	5·07	5·81	5·86
	19	4·07	5·29	5·63	5·35
	20	4·13	5·22	5·48	6·11

B. Interpenetrating Sample Means

Sample No.	Increments	Truck 1	Truck 2	Truck 3	Truck 4
1	1, 5, 9, 13, 17	4·302	5·148	5·442	5·796
2	2, 6, 10, 14, 18	4·292	5·230	5·676	5·580
3	3, 7, 11, 15, 19	4·276	5·164	5·630	5·790
4	4, 8, 12, 16, 20	4·314	5·096	5·466	5·930

usually legitimate to combine data from various stretches of belt, the assumption being that the within-zone variation is roughly constant under similar circumstances and for zones of similar size.

TABLE II

Analysis of variance: Ash contents of increments within and between zones

Source of Variation	Sum of Squares	D. of F.	Mean Square
Between zones ..	31·87	19	1·68
Within zones ..	2·55	60	$(\sigma_1^2 =)$ 0·0426

The standard error with which the mean of n increments, separated by the length of the zone, estimates the mean for the length of belt contained in n zones is then $\sigma_1/\sqrt{n}$, where σ_1 is estimated from the

within-zone variance. For the data of the example, if the mean for each truck is estimated from five increments, one increment from each zone, the standard error is $\sqrt{(0{\cdot}0426/5)} = 0{\cdot}092$ in the units in which ash content is measured.

METHOD B. The data for this method are in the lower part of Table I, and have been derived from the individual results by combining them according to the scheme shown in the table to give sixteen interpenetrating samples. In practice the increments for each sample would be bulked and a single determination made. The analysis of variance is performed in Table III, and it is inferred that the standard error with which the mean of a sample of n increments, taken at a distance apart equal to the zone length, estimates the mean for the whole length of belt is the σ_2 of Table III, which comes to 0·073.

TABLE III

Analysis of variance: Mean ash contents of interpenetrating samples

Source of Variation	Sum of Squares	D. of F.	Mean Square
Between trucks ..	5·42	3	1·81
Between samples within trucks ..	0·0635	12	(σ_2^2 =) 0·00529

The first method usually gives the most accurate figure for an experiment with a fixed number of increments. There may, however, be practical objections to analysing increments separately, and then the second method will be preferred. There is a theoretical objection to both methods in that the implicit assumption of zones differing in mean, but with random variation within them, is rather artificial; the zones are actually arbitrary in extent, and in general the mean will change continuously from one to another, not in discrete jumps. Comparison with the results given by the time series method given below, however, where these objections do not hold, suggests that in practice they are not very important. The practical objection to these methods only arises when it is desired to estimate the potential standard errors of means of a number of different systematic samples having different numbers of increments and different distances apart, with the ultimate aim of choosing the most suitable. For Method A it is possible to do this in some cases by regrouping the data into zones of different length and carrying out a new analysis of variance; but this is not very convenient, since each value of zone length examined requires an analysis all to itself. For Method B there does not appear to be any way of solving the problem, since the combined samples cannot be split up again, nor does the time series method appear to furnish a solution in the parallel case. It is, of course, important to bear in mind that the within-zone variation may be expected to decrease as the size of the zone decreases, so that the number of increments is not the only factor which directly influences the accuracy of the sample mean.

Time Series Method

The discussion of this method will be confined to the case where the experimental increments are analysed separately. The combined case may also be dealt with by this method but is of less interest, since here the time series treatment gives no particular practical advantage.

Suitable experimental material consists of sets of pairs of increments from the belt, the members of each pair being at a distance τ apart, τ varying from set to set. Assuming (as for the other methods) that variation over a zone of belt of fixed length is roughly the same for different zones, it is legitimate to combine data from various stretches of belt. One pair may have a member in common with another pair; the data should contain pairs with intervals varying from a size quite small compared with the anticipated ultimate distance between increments to a size at least a little larger. Data such as those in Table Ia would certainly be quite adequate (though much scrappier data would be equally adequate) for an anticipated final choice of distance between increments of about four minutes, as in the previous section. The mean squared difference of ash contents for pairs of increments separated by a given value of τ is then calculated for various values of τ. For example, taking $\tau = 2$ minutes, the data of Table Ia give for this mean square:

$$\frac{(4{\cdot}98 - 4{\cdot}80)^2 + (4{\cdot}58 - 4{\cdot}96)^2 + \ldots + (4{\cdot}13 - 3{\cdot}88)^2 + (4{\cdot}98 - 5{\cdot}11)^2 + \ldots + (5{\cdot}26 - 5{\cdot}42)^2 \ldots}{72}$$

$$= 0{\cdot}0934$$

This value is an estimate of the ordinate at $\tau = 2$ of a curve which may be described as the *mean squared difference curve*. It is useful to regard

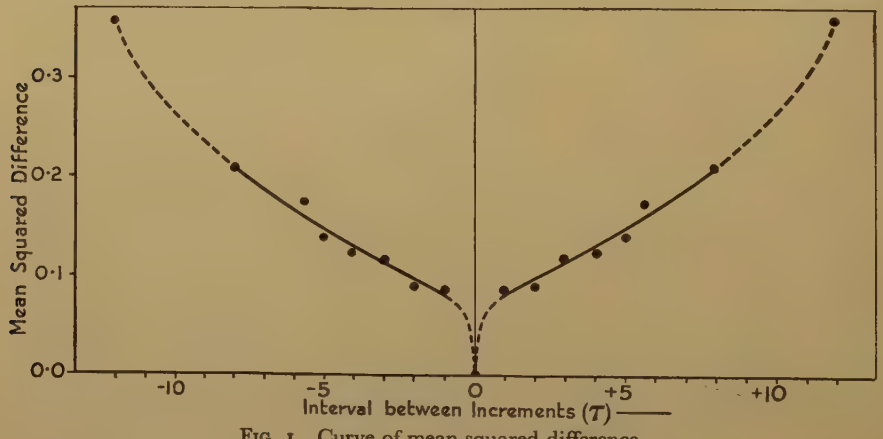


Fig. 1. Curve of mean squared difference.

this curve as being symmetrical about the origin of τ, taking the ordinate at $-\tau$ to be the same as that at τ. The ordinate is zero for $\tau = 0$, dipping sharply to zero at this value in most practical cases, and increases as τ increases in numerical value. The curve shows the extent

to which values of the variable at points τ units apart on the belt differ from one another. It is probably best fitted by eye; there is a possible numerical procedure, but it is somewhat complicated and laborious and is likely to yield little advantage in this application. For the example the mean squared difference curve is given in Fig. 1. In practice it will usually be found that only the central portion of the curve is needed, and it will therefore not be necessary to estimate ordinates for many values of τ. Those for the example are given in Table IV.

A second curve, unfortunately not so readily interpreted, is then deduced by multiplying the mean squared difference for each value of τ by the factor $1 - |\tau|/T$, where T is the length of the section of belt

TABLE IV

Mean squared differences: Ash contents of Table I taken at various intervals

Interval (min.)	Mean Squared Difference	Modified Mean Squared Difference ($T = 20$)
0	0	0
1	0·0850	0·0808
2	0·0934	0·0841
3	0·1185	0·1007
4	0·1284	0·1027
5	0·1391	0·1043
6	0·1725	0·1208
8	0·2074	0·1244
12	0·3585	0·1434

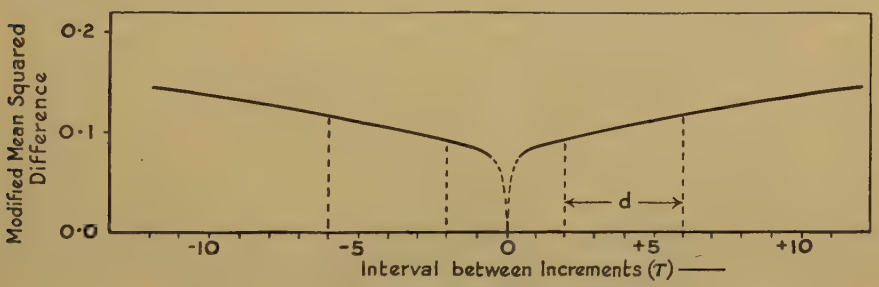


Fig. 2. Curve of modified mean squared difference.

which is to be sampled and $|\tau|$ is the value of τ irrespective of sign. We shall call this the modified mean squared difference curve, and denote its ordinate (taken as zero when $|\tau|$ exceeds T) by ϕ_τ. The curve for the example is illustrated in Fig. 2, the calculated ordinates being given in Table IV for $T = 20$ minutes. It could be determined directly if desired rather than from the curve of Fig. 1, but the curve of unmodified values is usually worth calculating because of the useful and fundamental information which it gives about the variation of the material on the belt.

If the routine sample consists of an increment taken from the same position (chosen at random) in each of n zones, each of length d, the

standard error $\sigma_{\bar{x}}$ of its mean is estimated as follows. The horizontal axis of the modified mean squared difference curve is divided into intervals of width d as indicated in Fig. 2. Next the average ordinate in the central interval is calculated, and possibly that in the adjacent intervals. The mid-ordinate is subtracted from the average ordinate, and the resulting figure added for all intervals. The total, when divided by $2n$, is then the square of the standard error of $\bar{x}$. In practice there will not be much curvature in the non-central intervals, and their contribution to the total will therefore be small, the average and mid-ordinates being much the same; the main contribution will come from the central interval, where the mid-ordinate is essentially zero, while the average ordinate is quite substantial, being about twice the size of the σ_1^2 of Table II. It is actually this fact which justifies the use of the first of the analysis of variance methods. The calculation of the average ordinates required above could be done either by counting squares under the graph and dividing by the interval width, or by some equivalent method. The graphical element in this analysis may seem a little unsatisfactory to a statistician who is accustomed to methods which are exclusively arithmetical, but the error involved will not be greater than that which arises from the inadequacy of the artificial assumption inherent in the analysis of variance procedure. In practice it will usually be most convenient to take the increment always at the centre of the zone. This is an error on the safe side, and will give a slightly more accurate mean than that resulting from strict application of the above procedure.

For the example, the standard error of $\bar{x}$ when $d = 4$ minutes and $n = 5$ is calculated as follows:

$$\sigma_{\bar{x}}^2 = \frac{1}{2 \times 5}\{\ldots + (0{\cdot}103 - 0{\cdot}103) + (0{\cdot}077 - 0) + (0{\cdot}103 - 0{\cdot}103) + \ldots\}$$

$$= 0{\cdot}0077$$

$$\text{i.e.} \qquad \sigma_{\bar{x}} = 0{\cdot}087$$

Similarly, the standard error for $d = 2\frac{1}{2}$ minutes and $n = 8$ proves to be 0·068. For this, the horizontal axis of the graph is divided into intervals of width $2\frac{1}{2}$, the rest of the procedure being as before.

In this discussion the possibility of a testing error in determining ash contents of individual increments has been disregarded. If such an error is present, its standard deviation being σ_0, the mean squared difference will dip sharply to the value σ_0^2 as τ decreases to zero, instead of dipping to zero. For purposes of application of the above procedure, however, the mean squared difference for τ exactly equal to zero must still be taken as zero.

Further Applications

These methods are equally applicable to such problems as sampling

atmospheric dust content in foundries at various times of day, or hardness at fixed intervals along steel rods, both of which involve within-zone variation and gradual changes in level from zone to zone. They may also be generalised to more than one dimension, e.g. to temperatures over a furnace wall.

MATHEMATICAL DERIVATION

In a series of observations in which the variational properties do not change with time (i.e. a stationary time series) the product-moment correlation coefficient between any observation and the observation an interval τ later is called an autocorrelation and is denoted by $\rho(\tau)$. For varying values of τ the function $\rho(\tau)$ is defined as the autocorrelation function, and it indicates how closely any term of the series resembles its successors and (for negative values of τ) its predecessors.

Suppose $x_1, x_2, \ldots, x_N$ is a set of observations forming a section of such a time series, each observation being separated from its successor by an interval w. If $N = nk$, this section may be divided into n equal zones, each containing k observations, as follows:

$$x_1, x_2, \ldots, x_k;\ x_{k+1}, x_{k+2}, \ldots, x_{2k};$$
$$\ldots;\ x_{(n-1)k+1}, x_{(n-1)k+2}, \ldots, x_{nk}$$

If the section is sampled systematically by taking an observation x_i from the first section, and observations $x_{k+i}, x_{2k+i}, \ldots$ from corresponding positions in the other sections, the sampling variance of the mean $\bar{x}$ of this sample is given by the following formula, due to Professor W. G. Cochran[1]:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(1 - \frac{1}{k}\right)\left[1 - \frac{2}{kn(k-1)}\sum_{u=1}^{kn-1}(kn-u)\rho(uw) + \frac{2k}{n(k-1)}\sum_{u=1}^{n-1}(n-u)\rho(kuw)\right] \quad \ldots\ldots(1)$$

A simple generalisation of this formula to the case of a continuous series $x(t)$ when n is large is given by Mr. M. H. Quenouille,[2] who gives:

$$\sigma_{\bar{x}}^2 \sim \frac{\sigma^2}{n}\left[1 - \frac{2}{d}\int_0^\infty \rho(\tau)d\tau + 2\sum_{u=1}^{\infty}\rho(du)\right] \quad \ldots\ldots(2)$$

where $\rho(\tau)$ is the autocorrelation function for the series and d the distance between the sampled points. An alternative generalisation, due to the author, which applies also when n is small, is obtained by taking the limit of (1) as k tends to infinity and carrying out some rearrangement. This results in (3):

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\sum_{-n}^{n}\left\{\left(1 - \frac{|ud|}{T}\right)\rho(ud) - \int_{(u-\frac{1}{2})d}^{(u+\frac{1}{2})d}\frac{1}{d}\left(1 - \frac{|\tau|}{T}\right)\rho(\tau)d\tau\right\} \quad \ldots\ldots(3)$$

where T is the length of the original section of series, the integrand

being taken as zero for $|\tau| > nd$. In view of the fact that for the series as a whole

$$\mathrm{E}(x_t - x_{t+\tau})^2 = 2\sigma^2\{1 - \rho(\tau)\} \qquad \ldots\ldots(4)$$

we can express formula (3) in terms of the function

$$\phi(\tau) = \left(1 - \frac{|\tau|}{T}\right)\mathrm{E}(x_t - x_{t+\tau})^2 \qquad |\tau| \leqslant T \qquad \ldots\ldots(5)$$

to obtain the new formula (6):

$$\sigma_{\bar{x}}^2 = \frac{1}{n}\sum_{-n}^{n}\left\{\int_{(u-\frac{1}{2})d}^{(u+\frac{1}{2})d}\frac{1}{2d}\phi(\tau)d\tau - \tfrac{1}{2}\phi(ud)\right\} \qquad \ldots\ldots(6)$$

If the reading is always taken at the centre of each section, similar formulae can be shown to apply, but the factor $(1 - |\tau|/T)$ in (3) should be replaced by the mathematically more complicated but approximately equivalent (tending to give a smaller standard error) factor:

$$\begin{aligned} &1 - \frac{2u}{n} + \frac{|\tau|}{T} && |\tau| \leqslant (n - \tfrac{1}{2})d \\ &-1 + \frac{|\tau|}{T} && (n - \tfrac{1}{2}) < |\tau| < nd \end{aligned} \qquad \ldots\ldots(7)$$

Normally, material on conveyor belts would be expected to show short-term variation due to segregation and other loading factors, daily and weekly cycles reflecting starting up and stopping down, and slow long-term variations due to changes in supply factors such as variation in the coal seam. The first would contribute to the mean squared difference a component of exponential type, increasing to a fixed limit as $|\tau|$ increases over the space of a truck or so; the second would be expected to contribute slow and smooth waves; and the third would contribute slow and smooth oscillations, which, though ultimately attaining stability, might take a long time to do so.

Over intervals as short as the zones described above, the second and third components would approximate fairly closely to straight lines, thus making negligible contributions to the terms of the summation in (6); further, because of its quasi-exponential nature and of the factor $(1 - |\tau|/T)$, the first component will usually only make substantial contributions to a few intervals in the centre of the range $-T = \tau = +T$. When the three components are taken together, therefore, only these central intervals will make substantial contributions to (6), and of these the most important will usually come from the central interval because of the cusp or sharp dip at $\tau = 0$.

Thus the method is applicable to very general stationary time series and does not depend on information about their long-term variation.

Usually error formulae associated with time series involve sums of long series of products of terms of $\rho(\tau)$, and are difficult to apply because of the difficulty of estimating the form of this function sufficiently accurately even for only moderately large values of $|\tau|$.

If n is large, Mr. Quenouille's formula applies, and the factor $1 - |\tau|/T$ may be neglected, the summation then being taken as infinite. This gives some theoretical simplicity but makes little difference to ease of application.

Thanks are due to Mr. J. Hebden of the National Coal Board, who raised this problem, and to Mr. J. F. Scott of the University of Sheffield, who constructed the example.

REFERENCES

[1] Cochran, W. G. (1946). 'The Relative Accuracy of Systematic and Stratified Random Samples from a Certain Class of Populations,' *Ann. Math. Statist.*, **17**, 164.

[2] Quenouille, M. H. (1949). 'Problems in Plane Sampling,' *ibid.*, **20**, 355.

QUESTIONS AND ANSWERS

Readers are invited to send questions to the Editor. He will have them answered by people competent to deal with the subjects raised.

Question

I have recently been trying to find data relative to various financial matters, and it seems that the sources of financial statistics are extremely hard to find. It appears that the Stock Exchange has no statistical department. Could you help me with the following points, please? (i) Where can I obtain particulars of shipbuilding in various parts of the world? (ii) Where can I obtain particulars of the number of securities quoted on the Stock Exchange and their nominal and market value? (iii) I believe that I have heard of a publication by the International Monetary Fund which is a useful source of information for international financial matters. Can you tell me the correct title of this, and how I may get copies?

Answer

As far as I know the Stock Exchange has no statistical department, but most of the larger member firms possess one. (i) Lloyd's Register of Shipping Returns, published quarterly at 71 Fenchurch Street, E.C.3, gives registrations of ships under construction in various parts of the world and other useful information on related subjects. For further information on sources of shipping statistics you may refer to Professor M. G. Kendall's *United Kingdom Merchant Shipping Statistics*, a pamphlet published at 2s. by the Royal Statistical Society. (ii) The Council of the Stock Exchange publish annually a summary of the number of stocks quoted, their nominal value and market valuation, marks received, etc. This is obtainable from the Clerk's Office, the Stock Exchange, Throgmorton Street, E.C.2. (iii) I think that the publication referred to in your third question is the *International Financial News Survey* or *International Financial Statistics*. These are published weekly and monthly respectively and may be obtained from the Office of Public Relations, International Monetary Fund, 1818 H Street N.W., Washington 25, D.C., U.S.A.

G. SALTER

Question

My firm contracts its market research to a firm of market research consultants who claim that they use a random sample which is representative of the whole country. I am led to believe that this method is better than the methods used by other agencies. I went to see this agency the other day and found out which towns they covered in the survey, and asked them to let me have data in respect of certain towns in the sample.

On examination of the data for these towns over a period of several months, it seems to me that there is a large error variation from month to month, and the sample in each town is extremely small. Would it not be better if they were to use fewer towns and take a larger sample in each town?

Answer

The question raises many interesting points, and some of them, if fully answered, would take us into quite heavy statistical theory. The question whether it would be better to select a few towns and to have a larger sample in each town raises a further question: 'Better for what purpose?' A larger sample for each town gives figures for any one town with greater accuracy, but such samples for a few towns do not give accurate estimates for the whole country. Samples are generally designed with a specific purpose in view. A sample designed to give national figures would be worthless for providing data for any one town, and similarly a sample designed to give a result for any one town would be worthless for providing national figures.

Some systems of sampling allow the interviewers to select their informants according to a number of characteristics; such systems are technically known as quota sampling, and for these there is no way of calculating the error. It may be large or small, you never can tell. In this respect random sampling is more valuable than quota sampling although it is more expensive, since a named individual has to be called upon and the interviewer must pay a second call if the informant is not available at the first visit; indeed, three calls may be necessary.

L. T. WILKINS

Question

Some of the figures which the cost office send to me every month are in the following form:

	This year to date	*Last year*
Output	3460	3396

The figure under the heading 'This year to date' is the total output since the first of January of the current calendar year, and that under 'Last year' is the total output for the corresponding period of the previous calendar year. Thus in the January report both figures are totals over one month, whereas in the December report they are totals over twelve. I would be grateful for your opinion on this form of presentation, and for a better alternative if there is one.

Answer

In answering this question four types of change in output rate have to be considered: (i) seasonal, e.g. a reduced output in December and August because of holidays; (ii) transitory, e.g. a reduction due to periods with abnormally high or low numbers of stoppages, not lasting very long; (iii) gradual, e.g. changes due to the cumulative effect of many little improvements in practice; and (iv) sudden but persistent, e.g. a change due to a major improvement in practice.

It is usually desirable that figures from the cost office should reflect types (iii) and (iv) as clearly as possible; types (i) and (ii) confuse the picture and are to be eliminated if possible. When output is totalled over several months, transitory changes which would noticeably affect output for the particular month in which they occur become much less conspicuous relative to the effect of gradual changes, since they tend to cancel each other out; such a total when corrected for the seasonal effect is therefore a good index of gradual changes. Unfortunately, however, it reacts only slowly to sudden but persistent changes, and it may be a few months before such a change affects it to a noticeable degree. On the other hand, a total of output over a single month, even if corrected for the seasonal effect, is very much at the mercy of transitory changes; and although sudden but persistent changes certainly reveal themselves with little time-lag, gradual changes may be swamped and consequently overlooked.

The only virtue of the method of presentation which you describe is that it automatically allows for seasonal changes. Unfortunately, the nature of the comparison which it presents changes as the year progresses. In January you may interpret a transitory change as giving signs of a gradual improvement. In December, on the other hand, your information, though reliable enough, is virtually six months out of date; if you have made a startling improvement in things in November, the effect may not be noticed until say January. It is therefore difficult to appreciate the significance of such data, for it is always changing. A good alternative would be some sort of compromise between the following two methods.

(i) Calculate the total for a complete year to date, e.g. November 1950 to October 1951, and compare it with a similar total for the previous year, November 1949 to October 1950. This will give a stable comparison, undistorted by seasonal effects, but with a time-lag. It is suitable for showing gradual improvements, but not sudden ones.

(ii) Determine the seasonal cycle from records of the previous few years—a suitable technique is described in any elementary textbook on statistics for economists, e.g. *Elementary Statistical Methods* by Dr. E. C. Rhodes—and use it to correct the current month's output and the previous month's output for the seasonal variation. This furnishes an index which will respond quickly to sudden changes, but is less successful than the other in revealing gradual changes.

As a compromise you might use, say, a six months' total, e.g. May 1951 to October 1951, and compare it with the previous six months' total, November 1950 to April 1951, making a correction for seasonal difference. The final choice depends on what effects you most want to detect.

G. H. JOWETT

NOTES AND COMMENTS

Businessmen so often complain of the burden of providing statistics for the Government that when one of them points out the usefulness of such information and asks for more, statisticians may perhaps be forgiven for calling attention to the fact. Sir Geoffrey Heyworth, a Past President of the Royal Statistical Society, devoted his 1951 annual address as chairman of Lever Brothers and Unilever Ltd. to the costs and efficiency of distribution to the consumer. In a discussion of the tasks of governments in promoting efficiency he stated: 'A second task for government is the collection and publication of statistics—of which the Census of Distribution at present being taken in the United Kingdom is one example. Its nearest parallel in Holland is the Census of Industry and Trade. The US, Canada, Australia, Italy, and the Irish Republic have published similar material. We have found that where such background material has been provided, studies of our own particular problems are more effective, and that where such statistics are available at regular intervals so as to enable changes in pattern to show themselves, they have still higher value. We think for a country to publish such statistics at intervals of five years is a practical objective. The fact that the US have published such statistics for a longer period than any other country probably explains at least in part their lead in the quantity and quality of more detailed and practical studies.' Sir Geoffrey then went on to elaborate his contention that the main contribution to efficiency must, of course, come from the trade itself. He mentioned the need for efficiency studies and market surveys both of which, we would add, require all the arts of the statistician.

★ ★ ★

The latest subject to come under the spell of statistics seems to be history. In a review of *The Incidence of the Emigration During the French Revolution* by Donald Greer, Mr. A. J. P. Taylor wrote in the *Manchester Guardian* of 29th June, 1951: 'Starting from figures often fragmentary and seemingly unpromising, Mr. Greer has dealt in a hundred pages of text and forty pages of statistical tables with a problem that would have taken a "literary" historian five hundred pages—and even then there would be no clear answer. Statistics open a new and most exciting door to history; and we ought to have books like this on all kinds of subjects. Though statistical history lacks literary grace, it brings increased knowledge and understanding in a different way; and ultimately the elegant writer can work its conclusions into a new synthesis. But what we need now are historians trained to use statistics—or, alternatively, statisticians with some understanding of history.'

★ ★ ★

A productivity team sponsored by the Anglo-American Council on Productivity has recently visited the USA for the purpose of studying American methods of inspection in the factory and had among its terms of reference the subject of quality control. A member of the team was Mr. E. D. van Rest, who among other things is chairman of the Industrial Applications Section of the Royal Statistical Society and a member of the editorial committee of this journal. Mr. van Rest was in the USA for six weeks with the team and made many visits to firms, mostly in the light engineering industries, whose inspection organisations were reported to have interesting features. He reports that the firms showed a variety of detailed practice, but what they had in common was their interest in the study of the subject of inspection itself, including both its management or organisational aspect and its detailed techniques. This was manifest also by the short courses run at many universities and by the activities of the many branches of the American Society for Quality Control. This society has experienced a tremendous growth in the last two or three years and now has a membership of over 4000. A Mid-West convention of fourteen sections of the society held in Chicago while the team was there attracted some 120 persons to a training course and many more to the convention itself.

A report on the findings of the team will be published in due course by the Anglo-American Council. This will contain references to the use of statistics in the manufacturing side of industry, and we hope in a later issue of *Applied Statistics* to publish an article by Mr. van Rest.

INDUSTRIAL APPLICATIONS SECTION OF THE ROYAL STATISTICAL SOCIETY

REPORT of the Section Committee for the Financial Year ended 31st December, 1950, and for the Sessional Year ended 30th June, 1951.

The Section Committee has the honour to submit its Sixth Annual Report.

Membership of Committees

The Section Committee, nominated by the local groups, by the 1949–50 Section Committee, and by the Council, consisted of:

W. A. Bennett and B. J. A. Martin (Birmingham)
E. C. Fieller and E. D. van Rest (London)
N. J. Squirrell and J. T. Richardson (North East)
A. H. Dodd and A. H. A. Wynn (Sheffield)
E. Lloyd and T. V. Starkey (South Wales)
D. J. Desmond, G. H. Jowett and Joan Keen (1949–50 Section Committee)
B. P. Dudding, W. J. Jennett and Philip Lyle (Council)

Mr. Jowett and Miss Keen were elected chairman and honorary secretary respectively by the committee.

The Local Groups elected the following committees:

Birmingham and District. W. A. Bennett (chairman), B. J. A. Martin (secretary), D. G. Beech, J. R. Cotterill, D. J. Desmond, J. A. Gilchrist, V. E. Gough, C. Wainwright, T. A. Yapp, and Elsie A. G. Knowles. Mr. M. Milbourn replaced Mr. Bennett as chairman during the session.

London. E. C. Fieller (chairman), E. D. van Rest (secretary), A. Blackwell, E. Day, C. G. A. Hall, R. H. S. Lesser, Philip Lyle, J. Maclean, D. R. Read, P. J. Stanley, and M. Whyte.

North East. N. J. Squirrell (chairman), H. Arnell (secretary), H. D. Gill, E. D. McCallum, J. B. Nadauld, W. J. Ross, and Doris Hyman.

Sheffield. A. H. A. Wynn (chairman), A. H. Dodd (secretary), P. J. Bovill, G. Fenton, J. Hebden, W. G. A. Jenkins, W. N. Jessop, and G. H. Jowett.

South Wales. T. V. Starkey (chairman), E. Lloyd (secretary), A. Beacham, L. Firth, T. J. Hayward, R. E. James, G. Loasby, L. Phillips, R. B. Rowson, and I. K. Walters.

Tees-side. H. Kenney (chairman), J. T. Richardson (secretary), F. L. Clark, H. Foster, A. B. Goggs, T. W. Hurst, T. C. Robinson, and W. Spendley.

Activities of Committees

The Section Committee met in full twice during the session, and subcommittees met on five occasions.

Business at these meetings included:

(*a*) Council's permission was obtained to publish the papers presented at the Sheffield Conference on Industrial Statistics (1950).

(*b*) Several discussions, some attended by representatives of the Study Section, were held, and recommendations made to the appropriate sub-committee on the contents of the proposed new journal, *Applied Statistics*. The Section Committee nominated Dr. O. L. Davies, Mr. G. H. Jowett and Mr. E. D. van Rest to serve on the joint editorial committee. It was suggested that *Applied Statistics* should report the business of the section.

(*c*) The British Institute of Management was asked and had agreed to include in its journal a note asking if any business firms would be prepared to take statistics students in their vacations.

(*d*) Mr. Desmond was nominated to represent the section at the Ninth International Management Congress and subsequently presented in Brussels a paper on 'Recent Developments in Quality Control.' This paper co-ordinated the findings of several countries.

(*e*) Explorations into the forming of a Glasgow group by Mr. J. Eaton of the Royal Technical College Mathematics Department were approved.

Membership

On 31st December, 1950, the roll of members was as follows, 1949 figures being shown in brackets:

Local Group	*Fellows*	*Non-Fellows*	*Total*
Birmingham and District ..	32 (29)	30 (42)	62 (71)
London	245 (258)	60 (48)	305 (306)
North East	8 (5)	17 (14)	25 (19)
Sheffield	16 (21)	53 (52)	69 (73)
South Wales	7 (5)	14 (16)	21 (21)
Tees-side	4 (3)	24 (26)	28 (29)
Totals	312 (321)	198 (198)	510 (519)

Finance

The accounts of the Section and of its Local Groups for the year 1950 (excluding expenditure incurred by the Society in relation to the work of the Section) are summarised below.

STATEMENT OF INCOME AND EXPENDITURE FOR 1950

Expenditure	£ s. d.	£ s. d.	*Income*	£ s. d.
Birmingham ..	16 3 5			
London ..	12 15 2		Grant from Council ..	109 0 0
North East ..	23 9 0			
Sheffield ..	15 17 0		Balance from 1949	34 9 6
South Wales ..	17 13 8			
Tees-side ..	8 19 0			
Section ..	25 17 2			
		120 14 5		
Carried forward to 1951 ..		22 15 1		
		£143 9 6		£143 9 6

Ordinary Meetings

Forty meetings have been held by local groups, comprising eight in Birmingham, seven each in London and Newcastle (North East), six in Sheffield, four in Billingham (Tees-side), three each in Crumlin (South Wales) and Norton (Tees-side), and one each in Newport and Cardiff (South Wales). Details of the subjects and speakers and some accounts based on local group reports are given below. Some of the meetings were held in conjunction with other societies, including the Institution of Production Engineers, the Institution of Works Managers, the Institute of Personnel Management, the Institution of Electrical Engineers, and the Sales Managers Association.

Special Activities

(*a*) Birmingham Productivity Exhibition. The Birmingham Group combined with the Institute of Engineering Inspection to organise the Statistical Quality Control Section of this exhibition. Over 8000 people attended, including 110 works parties, and at least 1000 separate enquiries were made regarding the application of statistics to the control of quality and as an aid to productivity.

(*b*) As a result of a questionnaire sent out to all members of the Birmingham Group to find what support it would have, the Department of Engineering Production at Birmingham University have organised a full applied statistics course for 1951–2.

(*c*) The London Group held a summer meeting at Guinness's Brewery at Park Royal, at which the work of a statistical section was described. This was followed by lunch and tea given by the directors, and it is understood that a sampling experiment was carried out on the product by anybody who felt competent to undertake it.

(*d*) In September 1950 the Section held its first conference at Sheffield University. More than 200 people attended, of whom 100 stayed in the university hostels. In addition to the meetings, visits to Carblox Ltd. and United Steel Companies Ltd., and a coach ride round the dales, were arranged.

The papers given on the first day were of a general nature and designed to encourage those who had not previously used statistical methods. The remainder endeavoured to illustrate new techniques, or novel applications of known techniques. They have been published by the Society.

The discussion on the proposed Industrial Applications Section of the International Statistical Institute led to the unanimous decision that an applied statistics journal was necessary, and the committee feels that the starting of *Applied Statistics* has been in some measure due to the activities of all those members of the section who, since 1945, have tried to spread the use of statistics among their colleagues in industry.

Papers filed in the Library

The following paper has now been added to the sixteen already filed in the society's library: *Statistics in Paint Technology*, by J. E. Mantell.

JOAN KEEN, *Honorary Secretary.*

E. D. VAN REST, *Chairman.*

14.12.51

APPENDIX

List and Reports of Meetings held by Local Groups

BIRMINGHAM AND DISTRICT GROUP

27th September, 1950. Statistical Indices: W. A. Bennett
25th October, 1950. Statistics and the Craftsman: Joan Keen
22nd November, 1950. Codification of Hospital Records: Professor L. Hogben
13th December, 1950. Statistics and Production Engineering: L. H. C. Tippett
28th February, 1951. The Interpretation of Inter-process Variability for Basically Different Products: V. E. Gough
21st March, 1951. The Use of Punched-card Equipment for Statistical Analysis: S. Brand
10th April, 1951. The Value of Statistical Methods of Control: P. Lyle
2nd May, 1951. The Rejection of Abnormal Experimental Results: Open discussion

LONDON GROUP

6th October, 1950. Statistical Problems of Personnel Management: B. Ungerson

Mr. Ungerson dealt with those problems of personnel management in which statistical methods could be of assistance: the handling and presentation of routine data, merit rating, job evaluation, and personnel selection.

In dealing with the first category he drew particular attention to data on labour turnover which, after correction for national trends, might provide a great deal more information to management than is usually required of it.

In merit rating schemes the statistician is required to say which of several tests gives the best results and which the best reproducibility between observers. Mr. Ungerson dealt in detail with the effect of the wording of the standard operational statements and with the check to be obtained by the distribution of marks. The analysis of merit ratings is often complicated by the need to use overlapping qualities to secure maximum reliability, and by the need to weight qualities so as to overcome the disproportionate effect of the more dispersed marks. The validity of selection procedures is subject to similar statistical considerations, but in this case the subsequent performance of the worker can be used as a check on the reliability of selection.

3rd November, 1950. Some Statistical Problems of the Paint Industry: J. E. Mantell

Mr. Mantell briefly surveyed the difficulties caused by the large number of items produced in different colours and in several sizes of containers, and by the batch nature of the process. He discussed in some detail an example from each of three fields: research, production, and sales.

His example from the research field was a derivation of a formula for transforming the colour assessment on one system to that on another. The formula could not be derived from theory and was derived by statistical examination of the relationship between the two assessments as made by 65 different panels of observers.

Mr. Mantell then described a production experiment designed to point to the most important factor influencing the filling of containers with the correct amount of material.

The sales problem dealt with size of batch and size of stock for minimum cost consistent with good service.

1st December, 1950. The Statistics of the Sieving Process: G. Herdan

Dr. Herdan introduced the subject by calling attention to the universality of the sieving process, which to the statistician was the equivalent of the grouping of observations. Though the process had been employed for many years, very little theory had been developed and little improvement made in the apparatus.

Standard sieves both in the US and UK are graded by aperture size. The usual quantities of interest are the weights of sieved material left between the sieves of a given series. Sieving is subject to error because its duration may affect the answer, and because the orientation of the particle passing through an aperture cannot be specified. The errors result in the disturbing fact that however prolonged the sieving process no agreement with expected values can be obtained except for very small loads; the fractions left between the sieves are always slightly higher than expectation, the effect of load being more serious with the smaller apertures.

The discussion made it evident that sieving is of importance in many industries. Professor Heywood spoke on the interpretation of the time curves and on the differences produced by different shapes of aperture.

8th February, 1951. Factor Analysis: Symposium

This meeting was held jointly with the Study Section.

2nd March, 1951. Some Production Problems and Planned Works Experiments: A. D. Grace

Mr. Grace gave a brief account of the functions of the Group Research Laboratories of Guest, Keen and Nettlefolds and of the wide range of articles manufactured by the many constituent companies; he then instanced large savings resulting from the use of quality control methods.

The two works experiments which the speaker described both arose from trouble during production. In one case planned tests on the actual production lasting one week, with corresponding tests on the ingoing material, revealed that more stringent control to a standard which could then be specified would be sufficient to cure the trouble.

The second example used an analysis of variance to trace the causes of deficiency in the electrical properties of some sheet steel. Allowance for the fact that parts of any one cast of steel might be rolled on different days and by different mills enabled the apparatus contributing to the aberration in properties to be isolated.

6th April, 1951. Subjective Judgment in Experimentation: G. Sears

Mr. Sears expressed the opinion that subjective judgment, properly handled, might be just as useful as measurement. Instead of being a method neglected except by such industries as food and dyestuffs, where there is no

other choice, it might well be used even where a measurement technique is available. He listed the various types of experiment which use subjective judgment, from the attempt to allot absolute qualities (as 'very good,' 'fair,' etc.) to the comparative test involving the simple comparison of a pair of samples for one attribute. He illustrated the applicability of the usual experimental designs by an example in which the smell and taste of three different oils were assessed by six assessors as good, fair, poor, and bad. One of the difficulties of applying a scale to such assessments is lack of homogeneity of variance, which can only be supplied by making a transformation, not necessarily the one that would make for the most efficient analysis.

A second example on the appearance of paint lacquers illustrated the use of the analysis of variance on the score obtained by partial paired comparisons. This analysis revealed an interesting interaction not superficially obvious.

The difficulties of extrapolating results obtained from a few assessors to those from a larger population, and difficulties arising from the indefiniteness of the boundary between subjective and objective assessment, were mentioned in the discussion.

4th May, 1951. Graphical Analysis of Variations as a Production Tool: E. A. G. Knowles and C. Roseman

This paper is essentially the same as one also presented by Miss Knowles and Miss Roseman at the Sheffield conference of the Industrial Applications Section in September 1950, and is reported in full in the proceedings of that conference.

NORTH EASTERN GROUP

18th October, 1950. The Measurement of Productivity: P. D. Vincent

Mr. Vincent stated that measurement of productivity could be applied to an individual factory or to an industry as a whole. Most of his work had been confined to the cotton industry, but the schemes used there could often be equally well applied to other types of industry.

After clarifying the differences between production and productivity Mr. Vincent limited his remarks to machine and human productivity, which are in general related. He defined productivity as the ratio of output to time, but pointed out that it was sometimes convenient to use other forms of measurement, e.g. efficiency (of machines) and standard minutes (operators).

Mr. Vincent then described methods of collecting the necessary data by questionnaires and field work, and dealt with the scope and analysis of this information. This description was excellently illustrated by means of slides showing graphs and diagrams based on data collected within the cotton industry.

7th November, 1950. Market Research: N. J. Squirrell

Introducing his subject to a joint meeting with the Incorporated Sales Managers Association, Mr. Squirrell said that he proposed to divide his address into two parts: firstly a broad outline of market research, what it does, and why; and secondly some graphical representations, with explanatory comments, of data collected.

By the collection of data and their intelligent analysis, the number of

decisions based on judgment before marketing a product could be substantially reduced; for example, the value of a product or the type of package preferred could be assessed by trials by users. Surveys could also be made on the size of the market, on comparisons with competitive or similar lines, and on the type of periodical or paper (dependent on the type of customer) it was desired to advertise in. The product having been marketed, its progress must be followed; and one way of doing this was to check with retailers and measure stock fluctuations at regular intervals.

Mr. Squirrell then showed graphs giving a typical distribution of a population, and explained how from this, sampling methods of obtaining information produced truly representative results.

13th December, 1950. Mechanical Recording and Computing Equipment: W. R. Dobbie and N. J. Squirrell

Mr. Dobbie confined his talk to a description and explanation of the uses and potentials of punched-card systems, using as his example the Hollerith system, though pointing out that all such methods were fundamentally similar.

A description followed of a typical card, with actual samples, and it was shown that a large accumulation of diverse data could be coded on to a perforated card, from which many different analyses could be made. Long and involved calculations could be completed with great economy of time and labour, and with absolute accuracy.

Mr. Dobbie then explained the sorting and extraction of particular data from records, and showed how the system could be used in sales analysis and market research, and in insurance and public health departments.

Mr. Squirrell then demonstrated the extraction of the standard deviation, using an electrically operated calculating machine.

17th January, 1951. Statistics and Industrial Health: H. Campbell

Mr. Campbell outlined the purpose of the Department of Industrial Health and explained that it was located in the north-east because this was an industral centre of varied large groups, e.g. coal-mining and shipbuilding, and that the main approach was not to examine the individual but a group or industry as a whole. Information was obtained from existing national data, from regional data, and from the analysis of the results of designed and other experiments. National data provided the bulk of the information, as records are available of every individual from birth to death, and birth-rates and death-rates in themselves are a measure of the health of the community. Statistics are also kept of the progress of the major diseases. Regional data could be obtained from surveys which were being made in individual factories and shipyards, and Mr. Campbell showed how he arrived at the sample sizes of people required to give positive information, citing instances where as a result of this work health had been improved in a region. Experiments were designed, in the main, to cover age-groups, sex, and occupation; and Mr. Campbell illustrated differences in these groups with the aid of slides of graphs.

He ended with a brief general survey of accident rates and the significance of variation of rates in different factories as compared with each other and with the national rate.

21st February, 1951. Statistics and Agricultural Research: F. Hunter and H. Aston

Mr. Hunter said that the increased use of statistics owed much to agricultural chemistry, one of the fundamentals of which was statistical development and analysis. He cited some of the different aspects of farming in which statistical knowledge was essential:

General farm organisation. An example showed that the cost of harvesting corn had been reduced from 9 to 3 man-days per acre.

The testing of crops, seeds being sent to different parts of the country to form random block experiments.

Animal husbandry, involving the compiling of records for progeny testing and of the egg yields of poultry.

Agriculture, botany, and insect control, and the effectiveness of different types of deterrents.

Engineering, and the value of different types of tools in different types of soils and locations.

Mr. Hunter gave several examples involving the use of Latin squares.

Mr. Aston then gave an example of the use of the Latin square in an industrial experiment.

21st March, 1951. A Comparison of Different Methods of Inspection: B. H. P. Rivett

Mr. Rivett stressed the importance to industry of a sound inspection scheme, and instanced the three main stations where inspection should take place:

Inspection of the raw material before processing to ensure that labour shall not be in vain.

Inspection of the material during processing to ensure that the process is satisfactory.

Inspection of the product.

Mr. Rivett said that the simplest and most obvious method of inspection was to test everything, provided there were sufficient time and labour available and sufficient confidence in large-scale inspection. Often, however, such large-scale inspection was impossible, as, for instance, in destructive testing; then the product must be sampled.

This led to an explanation of single and double sampling schemes, in particular the Dodge and Romig system, and thence to sequential sampling. Mr. Rivett illustrated graphically the operating characteristics and inspection power curves of these systems, explaining the methods used in trying to get a sampling scheme which most closely approximates to the ideal. He cited instances where 100 per cent. inspection could be less efficient than sampling inspection, e.g. where human fatigue supervened.

18th April, 1951. Statistics in Coal Technology: J. M. Runcie

Mr. Runcie outlined the scope of the scientific research carried out by the National Coal Board. The Scientific Department dealt with the quantity and quality of coal, supplied scientific data, made laboratory tests, and maintained a records office dealing with this information. As in each area there were many pits, and grades and seams of coal, a large sampling problem was involved. A measure of consistency of quality was also required, and

one of the factors used to control this was the ash content of the coals sampled, which had to be kept within control limits.

Statistical control was also used in washeries and cleaning plants. Float and sink analyses used straightforward quality control, the control level being the specific gravity, which could be adjusted to give the desired quality in the processed coal.

Mr. Runcie gave examples of the formulae used and described the methods of sampling. In the discussion which followed the sampling methods were severely criticised by several members.

SHEFFIELD GROUP

26th October, 1950. Operational Research on the Output of Open-hearth Furnaces. Control of Foundry Sand. Index of Operational Efficiency for Reheating Furnaces: W. N. Jessop

30th November, 1950. The Application of Statistics to Dust Problems in Coal Mines: J. G. Dawes

25th January, 1951. Statistics in Time Study: D. J. Desmond

22nd February, 1951. Statistics and Personnel Problems: G. H. Jowett

29th March, 1951. The Use of Statistical Method in the Development of Safer Mining Explosives: J. W. Gibson

26th April, 1951. Accident Statistics: A. H. A. Wynn

SOUTH WALES GROUP

3rd November, 1950. The Statistical Approach to Time Study: D. J. Desmond

8th December, 1950. Statistical Techniques in Personnel Selection: T. V. Starkey

2nd February, 1951. Incentives and the Young Worker: L. T. Wilkins

2nd March, 1951. Listener Research: R. Silvey

6th April, 1951. Statistical Methods in Machine Utilisation Studies: E. A. G. Knowles

TEES-SIDE SUB-GROUP

11th October, 1950. A Chemist's Views on Variability: K. H. V. French

8th November, 1950. Analysis of Production Costs: P. Lyle

13th December, 1950. The Number of Observations required in Experiments: G. P. Sillitto

10th January, 1951. Statistical Aids to Management: L. H. C. Tippett

14th February, 1951. Statistical Problems in Personnel Management: B. Ungerson

21st March, 1951. The Application of Probability and Statistical Method to Electricity Supply: R. B. Rowson

11th April, 1951. Some Aspects of Sickness Absence at Billingham: W. Spendley

STUDY SECTION OF THE ROYAL STATISTICAL SOCIETY

THE Study Section of the Society was established in 1928. After being in abeyance during the war it was revived in 1946, since when it has met regularly in London for the study and informal discussion of statistical subjects. The year 1950 saw a group of the Study Section successfully launched in Bristol, where it meets monthly during the Society's sessions in accommodation provided by the University of Bristol.

Meetings held during the 1950–1 session were as follows:

LONDON

12th October, 1950. Statistics in Criminology: M. Grunhut

9th November, 1950. Symposium on Problems of Business Forecasting: C. Nuttall, A. P. Zentler and J. Downie

14th December, 1950. Consumer Panels in Market Research: N. J. Squirrell and A. G. Cranch

11th January, 1951. Symposium on Statistical Problems in Distribution: A. W. McIntosh, B. T. Ramm and E. A. Lever

8th February, 1951. Applications of Factor Analysis: H. Eysenck, G. W. Scott-Blair, Max Hamilton, and Ardie Lubin—Joint meeting with the Industrial Applications Section

9th March, 1951. Studies in British Marriage and Fertility Data: P. R. Cox and E. Grebenik—Joint meeting with the Institute of Actuaries Students' Society

12th April, 1951. Some Field Problems in Family Budget Enquiries: Professor Ian Bowen

10th May, 1951. Annual General Meeting and Review of Literature.

BRISTOL

12th October, 1950. The Efficiency of Sampling Methods: G. Herdan

9th November, 1950. Some Developments in Market and Advertising Research in the USA: W. R. Foster

14th December, 1950. Prediction by Survey Methods: L. T. Wilkins

18th January, 1951. Some Aspects of Quality Control: E. W. Lott

14th February, 1951. Problems in the Evaluation of Quality: F. H. George

15th March, 1951. Problems in a Recent Sample Survey: B. D. Copland

The fifth post-war annual general meeting of the Study Section was held in London in June 1951, when the following committee was elected to serve during the 1951–2 session: Miss V. M. Cunliffe, Mr. C. A. Moser, Dr. Colin Nuttall, Mr. G. P. Salter, and Mr. E. Shankleman. The Council of the Society subsequently nominated two of its members, Miss J. I. Douglas and Mr. L. T. Wilkins, to represent it on the Section Committee. Meanwhile the group in Bristol had elected Mr. W. R. Foster, Dr. G. Herdan, Mr. H. C. Mackenzie, and Mr. G. F. Todd as its committee for the session. Mr. Foster

as chairman of the Bristol Group was subsequently co-opted to the Study Section Committee. The secretary of the group is Mr. Mackenzie.

Committee Meetings

Since its election the Study Section Committee has held two meetings. Business included the drafting of the current programme based as far as possible on recommendations and requests of members of the Section as intimated at the annual general meeting. Discussions also took place as to the contents of the proposed new journal *Applied Statistics*, the Section being represented on the Editorial Committee of this journal by Mr. C. A. Moser, Mr. E. Shankleman, and Mr. L. T. Wilkins. The committee has also had under consideration the formation of a special group to prepare a memorandum on the compilation of hospital statistics.

Stella V. Cunliffe, *Honorary Secretary*
Eric Shankleman, *Chairman*

REPORTS OF MEETINGS DURING AUTUMN 1951

LONDON

10th October, 1951. Hospital Statistics: F. D. Bushell

Mr. F. D. Bushell, Medical Records Officer at Westminster Hospital, London, read a paper entitled 'Some Statistics of a London Undergraduate Teaching Hospital covering the last Sixty Years.' While pointing out that many of the records were incomplete, he covered in detail the 'bed complement' and its distribution between various sections of the hospital, and explained that the distribution depended partly upon the specialists attached to the hospital. In-patients' admissions, out-patients' attendances, duration of stay, bed occupancy, and the cost of in-patients per week, were all discussed, and in addition he touched upon the various diseases, both fatal and otherwise, encountered at the hospital between 1890 and 1950. In the course of the discussion various members of the audience raised the question of the importance, under the new Health Scheme, of setting up some common method of recording hospital statistics. It was agreed that the Study Section might consider the matter further, with a view to making recommendations.

14th November, 1951. Market Research: A. Anson and T. Cauter

Mr. Anson, of Quaker Oats Ltd., read a paper describing what the sales manager expects to get when he employs market research specialists, and Mr. Tom Cauter, of British Market Research Ltd., discussed the question from the point of view of one who supplies market research. Mr. Anson said that sales managers relied on market research specialists to take the necessary precautions to ensure the reliability and statistical accuracy of information they provided. He felt that market research could be most useful to a firm that was proposing to enter a new market or to develop a new product; he described in detail the launching by his own firm of a product that had had considerable success abroad but had been rejected by salesmen in this country as being unsuitable for British markets. Market research had shown a demand for the product on the home market, and as a result his firm now holds 70% of the entire market for the product in this country. He wanted

assistance from market research specialists in developing programmes of selling (both at the wholesale and at the retail level) and advertising.

The sales manager used market research to make continuous and regular measurements of consumer buying. The consumer panel technique had been found to provide valuable information, but Mr. Anson had doubts about store audits because all possible outlets may not be represented. He had other doubts about the ability of market research to give a valid answer in such work as the checking of readership of advertisements, the testing of one package design against another, and the forecasting of how people would behave in the future under particular sets of circumstances. Mr. Anson stressed the importance of the results of market research being made available quickly, and he appealed to workers in this field to make their reports simple and easily understandable and to avoid bulky reports. He closed by pointing out that market research, although an invaluable aid to progressive sales management, did not provide a perfect answer to every sales problem.

Mr. Cauter did not believe that the whole function of market research was simply to supply facts at a competitive price. It was sometimes suggested that the supplier of market research should not be interested in the fate of the data he supplies, as the manufacturer would make what use of them he could. In these circumstances the research specialist anticipated a ready-made problem and gave a ready-made answer from his stock of suitable and well-tried techniques. Other manufacturers, on the other hand, regarded market research as a service; they called in the researcher, explained the problem, and asked for advice. They were not primarily concerned with the supply of facts, and the collection of facts was a secondary, if essential, interest; what they required was the best advice possible on the problem being tackled.

Mr. Cauter would like to see fact collection put on a more co-operative basis. He suggested that there would always be commercial organisations whose function it was to supply continuous data in specialist fields, but he felt that these would not be the true market research specialists of the future. He believed that to get the best out of market research, manufacturer and research consultant had to establish a continuous and close relationship, and this was facilitated by the formulation of an annual plan.

On the question as to whether manufacturers ought to establish their own research units or employ outside consultants Mr. Cauter felt the latter course was preferable unless the manufacturer in question had a very large business indeed. This applied in his view even when the manufacturer was concerned with highly technical material.

Mr. Cauter summarised what he considered to be the correct method of employment of a research organisation by a manufacturer: (i) to select the organisation for which the manufacturer had the greatest respect and then to take the research specialists into his confidence, (ii) to settle the question of remuneration on the basis of the number of interviews the field staff would make. With such an arrangement the consultant was able to draw on other research organisations for certain survey work. This was a help, as there was a tendency for some organisations to specialise in certain types of work.

Finally Mr. Cauter regretted that more had not been done to assist the development of new techniques. 'One of the outstanding features of market research today is the failure of new techniques to emerge. Every fundamental

method known and used today was known and used over a decade ago. Refinements, new applications, variations and larger scale there had been—but new methods, no!' He felt that manufacturers should support co-operative work into development of new methods and not leave the financing of this research exclusively to market research firms. He suggested that manufacturers' organisations in industries using market research should donate sums to one or other of the academic bodies investigating research techniques, and he drew attention to the work being done by the Division of Research Techniques at the London School of Economics, which had already produced data of great use to all commercial consultants.

The discussion that followed covered the link between advertising agents and market research practitioners, the question of bias introduced whenever human beings were employed as interviewers, and the use of psychologists, who, it was pointed out, themselves believed that they were very much still in the experimental stage in this field. The successful forecasting of expected demand was also considered, and it was believed that market research techniques had much to offer in this connection.

19th December, 1951. Pitfalls of Financial Statistics: Norman Crump

Mr. Crump, who is City Editor of the *Sunday Times*, said how glad he was to be back with the Study Section, particularly as he had been its first chairman when it was originally formed before the war.

He began by considering the pitfalls in what he called the 'bread and butter' side of the City Editor's job, namely stocks and shares. One was the question of the rise in the price of shares as reported in the newspapers. If a firm with two million £1 shares found that the price of their shares rose from 20s. to 30s. per share it would not be true to say that the total value of shares in the company had risen from two to three million pounds so that the holders of the shares were a million pounds better off. A further pitfall occurred when shares were quoted on one day at a pound and on the following day at 21s.; it would not be right for newspaper readers to believe that they would be making a shilling profit per share if they sold; brokerage, commission, etc., had to be deducted, and it could be that no profit at all would result.

Mr. Crump discussed the question of the value of capital. He pointed out that a nominal capital of £2,000,000 divided into £1 shares really meant that each £1 share entitled the holder to one two-millionth part of all property held by the company; 'the share' was a correct description and remained at one two-millionth part regardless of its market value. He discussed the distinction between the nominal and real capital of a company and the interpretation of balance sheets. On assets he warned that while their undervaluation might arise from a conservative attitude on the part of directors (which might be laudable) it was a very real pitfall when considering the actual financial position of the company and the real value of its shares.

By taking the assets at their real value, as far as this was possible, and by deducting all liabilities, including debentures, preference capital of all kinds, etc., one was left with the net assets of the company. These, divided by the number of shares outstanding, gave a picture of the real value of each share. Bonus shares were discussed and deprecated by Mr. Crump, who dealt in some detail with the advantages of the issue of shares for which no nominal

value is quoted, but which merely represent some proportion of the assets of the company.

The statistics of public finance, in particular the annual financial statement and the Treasury's weekly statement of revenue and expenditure, published in the *London Gazette*, were considered. The esoteric nature of these returns was emphasised; numerous pitfalls beset the user, and the definition of each item needed to be carefully examined. In particular it was suggested that it would be helpful if the subject of loans abroad were examined in greater detail.

BRISTOL

10th October, 1951. Problems in the Use of Consumer Panels: R. N. Wadsworth

Mr. Wadsworth suggested that market research techniques seldom produced unqualified results. The psychological factors influencing the behaviour of human beings still presented great problems.

A consumer panel was defined as a group of the same individuals submitted to a series of intermittent interviews or required to make a series of reports. Panels were compared with other types of market investigations such as *ad hoc* surveys, and retailers' sales and stock audits. The *ad hoc* survey appeared to have greater reliability in respect of details of actual consumption. Mr. Wadsworth was not certain that the consumer panel offered adequate evidence of the attitudes and opinions of consumers. The problem of 'conditioning' in consumer panels was particularly important.

A panel which involved the keeping of weekly diaries by its members was considered in detail. The choice of sample and the recruitment of the panel were described and the problem of replacing lapsed members considered.

All the results of consumer panels should be subjected to tests of accuracy, some of which Mr. Wadsworth described. These had revealed in the past that panels could substantially overstate true sales figures. Regional estimates of sales a head tended to be less satisfactory than other results. The solution was probably the use of larger samples if financially justifiable.

Mr. Wadsworth pointed out how useful the results of consumer panel enquiries could be in helping firms to know whether increased sales were due to buying by the trade or by consumers. The trend for various brands over a period of time was also a most important product of the survey, as was information on loyalty to a particular brand, the frequency of purchase, and the rate of consumption.

Mr. Wadsworth concluded that the consumer panel was a promising instrument of research, giving for some products more reliable information on purchasing than *ad hoc* surveys. It was a better method for studying changes in buying and spending habits and was capable of yielding data otherwise unobtainable. Being an expensive technique, the development of consumer panels depended on co-operation.

20th November, 1951. The Design of Experiments: P. K. Kenney

The principles developed by Professor R. A. Fisher for the design of experiments were briefly stated and their application illustrated by reference to two of the simplest designs: randomised blocks and Latin squares.

11th December, 1951. The Television Viewing Public: R. J. E. Silvey

Mr. Silvey, Director of Audience Research at the BBC, presented the methods and findings of a recent survey of the structure of the television public and the effects of television on leisure. A large sample of persons holding television licences and living within a fixed radius of the two television stations then in operation was interviewed. To offset the comparative novelty of television viewing the sample was made large enough to yield a reasonable number of persons who had had television sets for some years.

The information obtained related to the usual classifications, e.g. sex, age, income, size of family, education, etc. Records were obtained of the behaviour during one day of each individual in these households. The records included data relating to ordinary receiving sets, cinema visits, etc.

A control sample was obtained by interviewing persons in households without television sets in the immediate vicinity. The findings are due to be published in the April issue of the *BBC Quarterly*.

BOOK REVIEW

M. J. Moroney: *Facts from Figures.* Harmondsworth: Pelican Books, 1951. Pp. 472, 7¼in. × 4½in. 5s.

This book, besides being perhaps the most remarkable value in statistical books available at the present time, must be regarded as representing a landmark in statistics. When a subject is dealt with in a cheap popular edition it can only mean that the subject has 'come of age.'

The cover describes this book as 'a layman's introduction to statistics . . . with examples drawn from a wide variety of fields.' It seems doubtful, however, if many laymen will find their way unaided unless they are supported in their reading by a very keen interest in statistical methods. At times they will be out of their depth and at times a little impatient that the obvious is almost too fully explained. For example, the mean is discussed at some length, but the term 'parameter' appears many pages before it is defined on page 272.

Mr. Moroney admits that he is biased against several statistical applications. He is an industrial statistician and has made a very fine attempt at describing the methods and uses of statistics in industrial production, but his vigorous attack upon index numbers would be better omitted unless the argument could be supported by statistical reasoning and not the author's social theories. He has also some misgivings about medical statistics which would be better replaced by more objective comment. It is for time series, however, that Mr. Moroney reserves his most caustic comment, heading his chapter 'Time Series and Fortune Telling.' It is doubtless true that index numbers, time series in some forms, and some social statistics, are among the less respectable of statistical methods, but much recent development has taken place in these fields, about which it can only be assumed Mr. Moroney was unaware.

On the whole the book is not written in a light style. Only on occasions does the writer adopt a facetious or amusing style which seems to be reserved for medical, economic, and social statistics. Like all statistical textbooks, the final print is not without mistakes, but these are minor faults which will doubtless be corrected on reprint. On the whole the book is excellently produced.

Economic, medical and social statisticians reading this book, as I hope they will, can expect to be stabbed into justified annoyance from time to time, but there is a good chance of their gaining by this. A popular textbook which can find space for a simple introduction to discriminant functions, sequential sampling, confounding and replication, as well as most of the more generally known statistical methods, including a treatment of ranking methods which is fully up to date, is well worth much more than the price asked.

L. T. WILKINS

EDITORIAL

In an article on estimating the future population, printed elsewhere in this number, Mr. P. R. Cox refers to the distinction demographers draw between 'forecasts' and 'projections.' The argument runs roughly as follows: estimates are necessarily based on assumptions about the future, some of these assumptions almost inevitably turn out to be wrong, and so the estimates, regarded as forecasts, are unreliable although the process of making the estimates from the assumptions remains valid. It is disconcerting for the demographer to have the whole of his art and science rejected because of wrong forecasts or, what is worse, to have it classed with crystal-ball gazing; and so he uses the word 'projection' to describe the result of his processes of calculation.

This kind of experience is not confined to demographers. In the esteem of the relatively informed public, the whole technique of public-opinion polling only narrowly survived the fiasco of the forecasts of the results of the 1948 Presidential election in the USA; and all statisticians in real life risk losing face because their results have led to wrong advice being given and taken.

However, those who seek to apply statistics may not escape the difficulties of the situation by retiring into an ivory tower. Few people are interested in projections as illustrating the results of applying certain processes of calculation to certain assumed quantities; most readers, for example, will look at the projections given in Mr. Cox's tables to see what kind of thing is likely to happen. And the same applies to all statistical calculations. Men of affairs have daily to take decisions based on some forecast of the future, and if statisticians shrink from providing forecasts, using the best knowledge available to the best of their ability, they are not doing their job. Indeed, it is doubtful if statistical subjects will live for long unless they retain their relevance to practical affairs.

In an ideally departmentalised world, which would not be ideal in any Utopian sense, it would not be the function of the statistician to take responsibility for underlying practical assumptions nor to advise on courses of action, but in fact he often has to do so in some measure. Someone has to have the whole picture, statistical and practical, in mind, and if, as happens all too often, the man of affairs is unable or unwilling to apprehend the statistical picture, the statistician must apprehend the practical and advise accordingly. In so doing he must, of course, make clear the assumptions and qualifications as best he can, for he does not carry the ultimate responsibility for a decision. Plato thought that political salvation could come only through the rule of philosopher-kings. Perhaps a modern version of this view is that our affairs can only be really well managed by statistician-men of affairs.

ESTIMATING THE FUTURE POPULATION

PETER R. COX

It is important for many purposes not only to know the present size of the population and its composition, but also to make estimates of future trends. After dealing with methods of making such estimates Mr. Cox gives some of the projections of the population in Great Britain made for the Royal Commission on Population and discusses them in the light of such results as are available from the 1951 census.

WHY are calculations made about the future size of human populations? Many answers could be given to this question today, and most of them would refer to the assistance that such calculations can give in framing social policies and welfare plans. Originally and fundamentally, however, population estimates form part of man's attempt to increase his knowledge of the world and to reduce the uncertainty surrounding his life. Whether his guesses have proved right or wrong in the past, he continues to speculate about the numbers of his fellows, and forecasts for the future have the same attraction as—and often no greater degree of success than—prognostications about the weather, the economic outlook, or even the possibilities of interplanetary travel. The subject is of perennial interest.

Before the history, objects, and different methods, of population estimation are discussed it is desirable that the terminology of the subject should be explored. The word 'forecasting,' defined as previous contrivance or reckoning, is out of fashion in connection with population calculations, perhaps because it appears to imply an expectation of success that cannot be hoped for. 'Prediction' and 'prognostication' are words of the same kind and also have had associations with astrology and necromancy. 'Estimate,' representing a rough judgment or approximation, is used for short-period applications such as inter-censal numbers. The word 'projection' means throwing forward or delineation, and is often employed in connection with calculations of future population that are to be regarded only as illustrations of the results that would follow if certain tendencies persisted or if the most recent experience were to be repeated without change. Alternative bases are sometimes used in order to make a range of projections rather than a single calculation. The more reputable projections are accompanied with a written commentary in which a warning is given that it is not intended to try to forecast the actual future numbers but only to give a series of pictures as a general guide to those who are concerned about the way things may go. These cautionary statements tend to be

overlooked, however, by laymen who devote only a short time to the perusal of the principal features of the projections.

Historical Survey

The extent of public preoccupation with the size of the population has varied from time to time in accordance with religious beliefs and historical situations, but there have always been practical reasons for counts and surveys. These may be seen from old records such as the early chapters of the Bible. Censuses were taken for taxation purposes and in order to raise levies for military service. Abraham was promised that his seed would multiply greatly, and this no doubt seemed very desirable, for it implied power and strength in battle.

Population projection, however, cannot be said to have begun until modern times. The regular and comprehensive censuses that commenced to be taken in the eighteenth and nineteenth centuries were necessary before guesses for the future could be transformed into estimates. In many of the countries concerned the population was growing rapidly. In the New World, especially in the USA, this expansion was of interest in relation to the development of the untapped resources that were waiting to be used. In Europe there did not seem to be room for expansion, and prophecies were made of starvation, pestilence, and vice, as the result of overcrowding. Malthus championed the idea that the potentialities for human reproduction were greater than those for increasing the food supply. However, industrial advances and the opportunities for emigration helped to remove such checks on population growth and confidence in the future was restored. In such an atmosphere prognostications of population do not seem to attract the attention that they receive when they are gloomy and full of foreboding.

Towards the end of the nineteenth century there began a progressive reduction in the size of the family. As mortality was also falling, the numbers of children remained for many years sufficiently large to provide the prospect of an increasing population, although at a diminished rate of growth; but eventually it seemed to some people that not only were too few children being born for replacement purposes but the shortage was serious and permanent and might become even more grave. Some population projections were published showing the long-term results of such apparent tendencies and, mainly because of differences of opinion on the extent of the insufficiency of births, the appearance of several further sets of calculations was stimulated. So great was the public concern that in 1938 an Act was passed in Parliament providing for the collection of more adequate fertility statistics, and in 1944 a Royal Commission was appointed which, in the course of its extensive investigations, made some further projections.

Since the end of the war in 1945 the study of population has been extended from the national to the international scale. The United Nations Organisation has a demographic staff. The principal

preoccupations are now the changing potentialities of the Great Powers and the increase in world population in relation to the food supply. Neither of these lines of thought is new but, as shown above, they both spring from the application in modern circumstances of ideas which have had much importance in the past.

Besides these main objects of attention in making projections there are many other reasons for wanting to know the trend of population in particular countries, and there is much demand for information on the subject, especially in America. The period of time for which estimates are needed for specific purposes is frequently shorter than when the wider question of general demographic trends is under consideration, and often the range of ages is restricted in some way. Among calculations that are more limited in scope may be mentioned the assessment of the size and distribution of the population at various points of time between censuses in order, for instance, to be able to calculate birth rates and death rates in the administrative areas. The Registrar-General for England and Wales publishes annual estimates of the numbers of men and women in quinary age-groups in this country, and recently has subdivided these by marital status.

Government departments and industry alike have their requirements relating to future numbers. For example, an article in the *Financial Times*[1] not long ago dealt with the subject of supply and demand in the perambulator and baby-carriage trade; the correlation between the sales of these articles and the numbers of births in recent years was shown to be high. There is an obvious connection between the size of the future population at the appropriate ages and plans for the construction of schools, the training of teachers, the recruitment of labour, the call-up for military service, the provision of houses, and the financing of retirement pensions.

Calculations of the future population for terms of years varying from one to thirty, or even more, have been a feature of social insurance planning ever since the first Old Age Pension was granted in 1908. Short-term estimates are made for budgetary purposes, and projections of longer range aim at illustrating the broader financial trends. In this field the preparation of cost estimates covering a period of years is desirable when plans for schemes or for the revision of schemes are under consideration. This is because often the main financial liabilities are deferred and the initial costs are misleadingly low. The basis of such cost estimates is a projection, in whole or part, of the population of the country concerned. Calculations of this kind give a sufficiently good picture of the course of events, although it is appreciated that close accuracy cannot reasonably be expected. Of course, alternative projections can be made in an attempt to show the limits within which the probable cost will fall, but frequently preference is expressed for a single series of figures for future years because ranges are apt to be a source of confusion to the administrator.

The various methods of population estimation and projection, and

the likelihood of their proving accurate, are associated with the nature of the work undertaken and the range of ages involved. It is proposed to speak briefly of three classes of calculation namely: short-term estimates, curve-fitting to express the trend of population, and normal methods of projection for general purposes.

The estimates published by the Registrar-General each year are made by taking the latest census figures classified by sex, age, and marital status, and adjusting them by means of the data for birth, marriage, and death registrations, during the period that has elapsed since the census. Although some difficulty is experienced through misstatements and inaccuracies in the census and registration material, and complete information about local migration may be lacking, these estimates are satisfactory as denominators for calculating birth rates and death rates and for many other uses. The degree of accuracy attained in the subdivisions by marital status has yet to be assessed in the light of the results of the 1951 Census.

Where a system of national registration is in force under which the notification of essential details of population movement is required, estimates of this kind have no longer any prognostic quality except in regard to such items of information as are not provided by the system. It often happens, however, that registration is incomplete or otherwise unsatisfactory, and there is the danger that the system may be brought to an end, as has recently happened in this country. A variety of methods may be used, including the study of trend curves, in order to arrive at the estimated size and distribution of the population. One of the strangest is perhaps that advocated in a recent article in the *Journal of the American Statistical Association*[2]; it was suggested that the birth rates and death rates of certain areas should be extrapolated in the light of national and other trends and then divided into the current figures of birth and death registrations respectively so as to obtain up-to-date estimates of the population in these areas.

The Registrar-General for England and Wales recently issued some short-term population forecasts showing the expected numbers of men and women two or three years ahead. In the absence of any major disturbance such as a war, a high degree of accuracy may well be achieved in such work. The technical problems encountered are different from those belonging to projections for longer periods.

Population Curves

The fitting of mathematical curves to data for the total size of the population has the attraction of simplicity in comparison with the more laborious calculations involved in alternative processes but is less informative on matters such as age-distribution. Success in forecasting by this method has been elusive, and the interest lies mainly in studying the behaviour of the numbers in the past. The type of curve chosen should, however, be in harmony with correct general notions of what a population may have been and may become.

If the rate of reproduction remains constant, the size of the population increases or decreases ultimately in geometric progression. The curve for the population P_t at time t associated with this concept is the logarithmic

$$P_t = Ce^{rt}$$

where C and r are constants, and this has frequently been referred to when discussing the trend of population size. Malthus postulated a geometric rate of increase to the population but only an arithmetic rate of increase to the food supply in order to illustrate his idea that the potentialities for human reproduction were such as to outstrip production of the necessities of life. The mathematics of the stable population has been founded on the same curve.

It is clear, however, that there must be some limit to the numbers that can inhabit a given space; an infinite population is hardly possible. The rate of growth must therefore diminish, and instead of the differential equation

$$\frac{1}{P_t}\frac{dP_t}{dt} = r$$

that underlies the logarithmic curve one may expect some modification such as

$$\frac{1}{P_t}\frac{dP_t}{dt} = r - f(P_t)$$

The simplest form for $f(P_t)$ is kP_t (where k is constant), and if this is used the so-called 'logistic' curve is obtained, namely

$$P_t = (Ce^{-rt} + k/r)^{-1}$$

that was originally developed by Verhulst and later studied by Pearl and Reed in America. The upper limit for P_t is r/k.

An interesting series of experiments was made by Pearl and Reed[3] in which the rapidly-breeding fruit-fly *Drosophila* was used. The numbers of these flies existing, with an adequate food supply, in a confined space were measured at regular intervals, and it was found that they reached an upper level beyond which they did not pass. Some evidence was thus found in favour of the logistic or a similar form of population growth.

The conditions of human life are more complex, and it is questionable whether one logistic would adequately express the trend of the numbers of men and women in any country over a long period if the actual data were available for such a test. Although success has been achieved in fitting the logistic to the populations of England and Wales, France, and America, for much of the period since censuses began, the sections of the curve that have been used are only partial. Yule[4] drew a logistic through points on a chart representing the data for this country for the years 1801 to 1911 and obtained a not unreasonable fit; the upper limit to this trend was assessed by him as being 90–100 millions

according to the precise method adopted. There is nothing inherently impossible about such an ultimate size, but since 1911 the enumerated populations have been less than those on the curve, and projections that take the various aspects of the current population situation into account show that there is little chance of the recorded numbers rejoining the curve within the next few decades.

A comparison of the actual total population of England and Wales with values of the logistic function as fitted by Yule is given in Table I. For the period up to and including 1921 the figures are Yule's; those for later years have been calculated for present purposes. Of all the alternative projections made by the Statistics Committee of the Royal Commission on Population, the one that gave the highest rate of growth in future numbers suggested an increase by 1971 of no more than 10 per cent. over the present level of nearly 44 millions. The logistic curve produces, however, an addition of about 15 per cent. to the figure of 51 millions for 1951. The ultimate population on the curve is 97 millions, whereas the highest figure quoted by the Statistics Committee is 61 millions in the year 2047—the lowest being 30 millions.

TABLE I

Population of England and Wales in millions: actual values and values calculated from logistic function

Year	Actual	Logistic Value	Difference
1801	8·89	8·93	+ 0·04
1811	10·16	10·32	+ 0·16
1821	12·00	11·89	− 0·11
1831	13·90	13·66	− 0·24
1841	15·91	15·65	− 0·26
1851	17·93	17·87	− 0·06
1861	20·07	20·33	+ 0·26
1871	22·71	23·02	+ 0·31
1881	25·97	25·95	− 0·02
1891	29·00	29·11	+ 0·11
1901	32·53	32·47	− 0·06
1911	36·07	36·02	− 0·05
1921	37·89	39·72	− 0·17
1931	39·95	43·53	+ 3·58
1941	41·75*	47·41	+ 5·66
1951	43·74	51·30	+ 7·56

* Estimated.

If it is supposed that the total population size changes in accordance with some mathematical function it must be admitted that there is a definite relationship between the courses of the constituent items of movement, namely births, deaths, and migration. These and the corresponding rates per thousand population cannot be considered as independent but as subject to a constraint. Given a certain expected development in the death rate and certain assumptions about migration, for instance, the trend of the birth rate is already determined and cannot be considered as an independent entity. Various methods of

analysis of mortality, fertility, and migration, have not suggested, however, that there is any such master connection between them.

Shortly before the time when Yule published the results of his researches with the logistic curve, Bowley[5] had made some calculations bearing on the future size of the population of Great Britain. In these he employed as bases the mortality rates of England and Wales in the years 1910–12 and the average annual numbers of births in Great Britain during 1921–3. On the assumption that these would continue unchanged, he examined what the course of the total population size would be for some decades to come. Although each author disclaimed that his work was a prophecy for the future, there was a natural tendency to compare and contrast the two methods in this connection. Yule explained the shortfall in the population of 1921, as compared with the logistic curve, as being due to the war of 1914–18; apart from this, the logistic still appeared plausible at that time as a curve representing population growth. He regarded Bowley's estimate for 1971 of 49 millions in Great Britain as too low. From our standpoint today, however, we see that this is by no means an unreasonable figure.

Bowley was able to point out, by way of reply, that the logistic was not the only curve that could be fitted with success to the data for the populations of various countries. The cumulative distribution function of the normal curve of error, or even a parabola, was capable of giving as successful a representation of the trend of the census numbers in England and Wales. The ultimate level, if there was one, might well differ markedly from that given by the logistic.

Since 1925 it has become clear that suitable formulae in connection with changes in the total size of the population have not been developed and that those tried thirty years ago have not proved successful. No doubt today it would not be impossible to obtain plausible results by joining two or more logistics together or by using more tortuous curves, but there could not be much confidence in their predictive power unless special regard was had in choosing them to the present age-structure of the population and the current tendencies of birth rates and death rates.

Population Projections

Attention to such points is fundamental in the methods of population projection that have found favour in recent years. If some assumption is made about the rates of mortality that will be experienced during the period to be covered, for each age of the population, it is readily possible to calculate survival factors and so ascertain the numbers of people among the present population who will still be living at any time or at any age in the future. In order to complete the picture it is necessary to add in an estimate of the number of births in each future year and then to estimate the proportions of them who will survive to various ages. An illustration of the technique of projection in this fashion has been given elsewhere[6] and need not be explained here.

Considerable interest attaches to the different methods used at various times for assessing the mortality rates for use in projections and for arriving at the numbers of future births. These also have been described elsewhere[7] and only a few general remarks are appropriate in the present connection.

The success or failure of a population projection depends on the stability of mortality rates, fertility rates, and migratory movements, or of their trends as deduced from such data as are available to the demographer. The downward course of the death rates in many countries during the past few decades and the prospect of further falls provide a ready general guide for the assessment of mortality bases for projections. The speed of progress in prolonging life is, however, variable, and it is not possible to foresee precisely what new medical treatments will be devised or the effect that they will have on longevity at the various ages. Thus complete success in mortality forecasting has not been achieved; in general there has been a reluctance to assume, in advance, rates of decline in mortality as rapid as those which have subsequently been experienced. Such a form of caution is unwarranted in estimates where an error is as important on one side of the truth as on the other.

During the last twenty years there has been intense interest in the course of the birth rate and marriage rate and in intrinsic fertility, analysed in various ways. It has, however, been a period much disturbed by the war, and the advent of new data and fresh ideas has changed the methods of approach to this subject. It is thus difficult to assess the degree of accuracy that is likely to be obtained in estimating the numbers of births in future projections. Confidence is not very high at the present time in this country, and there are doubters even in America, following repeated failures to foresee developments in family building.

This want of confidence finds expression in the simultaneous publication of series of projections on alternative bases rather than a single projection, and the range of the results thus obtained may be wide. No precise statistical significance can be attached, however, to the upper and lower limits of the ranges, nor indeed are these necessarily chosen to represent extreme assumptions. Unless the demographer selects an optimum forecast from among his various calculations his work does no more than illustrate the results that would follow if certain tendencies persisted, upon the likelihood of which he does not comment, and show which trends deserve encouragement and which should, if possible, be resisted or altered.

It is of interest at this point to have some indication of the extent of the discrepancies that have arisen between the population of this country in 1951 and the corresponding projected numbers that have been published at various times within the last twenty years, excluding the Royal Commission figures. In general the projected totals were too low, the shortfall ranging from half a million to about four millions according to the projector and to the assumption made, that is, from

one to ten per cent. of the size of the population. Most of this discrepancy arose at the youngest and at the oldest ages, the biggest deviation of actual numbers from projected being about 2 per cent. at 15–64, 10 per cent. at 65 and over, and 35 per cent. at 0–14. These figures illustrate the relative difficulties of dealing with the various stages of life in projections made over periods of ten to twenty years.

Although in some instances alternative projections were quoted by the author, the population has thus exceeded even the most favourable expectations that he chose to express. These affairs do not appear to be ordered any better in other countries. For instance, according to Lorimer[8] a projection of the population of the Soviet Union that was published in 1934 in the names of the demographers Novoselsky and Payevsky gave the estimated population in 1939 as 191 millions, whereas the census of that year revealed in fact a total of only 170 millions. Of the total discrepancy of 21 millions, fifteen millions was attributable to the first twelves ages of life.

Royal Commission Projections

In their report[9] the Royal Commission on Population appear to have taken as axiomatic the need for the preparation of a comprehensive series of projections in order to consider the 'probable consequences' of the 'present population trends,' as they were requested to do by their terms of reference. Having finished their consideration of the current level of births in relation to the numbers needed for replacement purposes, they discussed at some length the 'future possibilities' conditioned by the 'existing population.' They used two alternative mortality bases, five sets of marriage rates, five fertility bases, and a range of possible types of migratory movement. Sixteen projections are shown in the Statistics Committee Papers[10] to have been completed; they may be regarded as consisting of one 'master' set of assumptions with others successively examining the effects of variations from the master set in each separate basis singly. Of these only a limited number were mentioned in the main report, namely those with a net migration of nil, with falling mortality, with intermediate marriages, and with three possible fertility levels. The full ranges of alternatives and the limited selection may be judged according to the taste of the observer; for instance it may be argued that the use of the mortality of 1942–4 as a basis produces a series of death rates that are too high for the future.

Now that the technique of illustrating demographic trends by means of projections has become established it is likely to be repeatedly necessary for fresh calculations to be made in order to correct the errors revealed by time in previous series. Already new population forecasts are demanded at frequent intervals in America by users whose faith has not been shaken by unfavourable past experience. In Great Britain, however, the monumental work of the Royal Commission, completed within the last five years, still holds the field, and it is of considerable interest to examine the results of these projections.

Some of the figures given by the Statistics Committee are set out in Tables II and III. Although for illustrative purposes the Royal Commission's projections were carried forward for as much as a hundred years, which is too far for the results to be credible, the main attention in the Report was devoted to the years 1962 and 1977 and it is to these years that the accompanying tables also refer. For these dates the central group of projections gives a picture different from that presented by pre-war estimates and is thus a valuable corrective to the former over-gloomy views. The new calculations reflect the more favourable wartime developments and have thus fulfilled a useful function apart from the question whether they will remain accurate as a measure of the population of Great Britain as time passes.

It may be seen from the tables that, on any reasonable assumption, the size of the total population is unlikely to alter much in the next twenty-five years; whether or not a peak is reached and passed depends on the projection chosen. The prospects for later periods must depend to a considerable extent upon the number of children under the age of 15 in the country by the year 1977. This depends more on the assumptions made, and is seen to vary from 8·1 millions to 11·5 millions, or by as much as 40 per cent. The elderly—of ages 65 and over—are likely on any projection to grow in number; from 5·0 millions in 1947 they may be expected to increase to 7·2, 7·9, or 8·1 millions, i.e. to 15 or 16 per cent. of the total population. At the principal working ages there is little room for difference of opinion on what will happen in the next few decades; the change from now to 1977 is less marked.

The tables are arranged so as to exhibit the variations in projected populations consequent upon changes of particular bases. In Table II there are five columns, of which the first two show the effects of an alteration of assumption regarding marriage. It may appear surprising that male and female marriage rates for the same period of time should produce different figures, but it is a fact that whereas the numbers of men and women marrying are the same each year those available for marriage are not. The sexes are born in unequal numbers and are affected to an unequal extent by mortality, by migratory movements, and by wars. Thus their marriage rates are not the same; male rates provide a high marriage basis and female rates provide a low marriage basis. The total population by 1977 depends to the extent of about a million persons upon the choice of assumption in this connection. It is, of course, the numbers of children that are most influenced by the difference of basis.

The third and fourth columns in Table II differ only in regard to the assumption of a constant or declining mortality, the discrepancy in the 1977 population being in this case rather more than a million and arising chiefly at the older ages. The fourth and fifth columns show how an annual outward movement of 100,000 persons would bring about a reduction, chiefly in middle life, in comparison with an absence of net migration.

TABLE II

Projections of the population of Great Britain, in millions, made by the Statistics Committee of the Royal Commission on Population. The effects of variations in the marriage, mortality, and migration bases

Projection No. ..	1	2	6	8	12
Mortality basis ..	Constant	Constant	Constant	Falling	Falling
Fertility basis ..	1935–8 rates	1935–8 rates	5 per cent. above 1935–8	5 per cent. above 1935–8	5 per cent. above 1935–8
Marriage basis ..	Female marriage rates of 1942–7	Male marriage rates of 1942–7	Inter-mediate	Inter-mediate	Inter-mediate
Net migration basis	None	None	None	None	Net emigration 100,000 a year
Under age 15 1947	10·3	10·3	10·3	10·3	10·3
1962	9·7	10·1	10·2	10·4	10·0
1977	8·7	9·2	9·4	9·8	9·0
Ages 15–64 1947	32·8	32·8	32·8	32·8	32·8
1962	33·1	33·1	33·1	33·5	32·2
1977	31·2	31·6	31·7	32·7	30·0
Ages 65 and over 1947	5·0	5·0	5·0	5·0	5·0
1962	6·0	6·0	6·0	6·3	6·2
1977	7·2	7·2	7·2	8·1	7·9
All ages 1947	48·2	48·2	48·2	48·2	48·2
1962	48·9	49·2	49·4	50·2	48·4
1977	47·2	48·0	48·4	50·7	46·9

TABLE III

Projections of the population of Great Britain, in millions, made by the Statistics Committee of the Royal Commission on Population. The effects of variations in the fertility basis. A falling mortality, an intermediate marriage basis, and no net migration are assumed

Projection No.	7	9	10	11
Fertility basis	1935–8 rates	Exact replacement	Falling	Rising
Under age 15 .. 1947	10·3	10·3	10·3	10·3
1962	10·1	10·9	10·0	10·8
1977	9·3	10·5	8·1	11·5
Ages 15–64 1947	32·8	32·8	32·8	32·8
1962	33·5	33·5	33·5	33·5
1977	32·4	33·2	32·3	33·2
Ages 65 and over .. 1947	5·0	5·0	5·0	5·0
1962	6·3	6·3	6·3	6·3
1977	8·1	8·1	8·1	8·1
All ages 1947	48·2	48·2	48·2	48·2
1962	49·8	50·6	49·7	50·6
1977	49·9	51·8	48·5	52·8

Table III shows the results of projections on four alternative fertility bases. By 1977 the influence of a change of assumption is felt only a little beyond the first fifteen years of life. These are nevertheless an important pointer to the general trend of population size in periods further ahead.

The 1951 census enumeration has revealed, according to the preliminary reports, a total population of 43·74 millions in England and Wales and of 5·10 millions in Scotland. The age-distributions of these numbers are not known at the time of writing, but the sample census data soon to be published should enable them to be accurately estimated before the end of this year. For 1947 the total population of Great Britain may be taken as 48·19 millions, and interpolation from the projected figures set out on page 241 of the Statistics Committee Papers gives an estimated population at mid-1951 varying from 48·65 to 49·50 millions—the latter assumes heavy immigration. An independent investigator may wish to try to judge which of the sixteen alternatives quoted by the Statistics Committee have so far been the most representative and the extent to which they appear to be relevant for the future. For this purpose it is desirable to examine the principal developments in the population since 1947 in more detail.

Recent numbers of births and marriages in Great Britain have been as shown in Table IV.

TABLE IV

Year	Live Births	Marriages of Spinsters (ages under 45)
1948	877,000	378,000
1949	827,000	357,000
1950	785,000	341,000
1951	770,000 (approx.)	340,000 (approx.)

The figures given by the Statistics Committee relate to the average annual numbers between mid-1947 and mid-1952 and are set out on pages 251–3 of the Papers. The range of the numbers of marriages of spinsters under age 45 is from 277,000 to 337,000 a year in the first six projections, and all of them are thus exceeded by the actual numbers. In probable association with this excess, the actual numbers of births compare favourably with all the projections, but cannot be regarded as falling outside their range.

In general, death rates have been, in the last five years, below the assumed 'falling' series, especially at the younger ages. The effect of this lighter mortality upon the number of deaths is, however, not great, and over the period 1947–51 they fell short of the expected average of 580,000 a year by only some 2 per cent. There was a net emigration from this country of over 50,000 a year during 1947–9.

There is thus no projection that has been completely successful in

representing population developments in this country—such an outcome was not to be expected. Nevertheless there are several of the projections that show for 1951 a population little different from that enumerated, at least in total. Most of them, including No. 8, No. 9, and No. 10, that were quoted in the Royal Commission's Report, give results that are somewhat too high because of the assumption of no net emigration, but a correction may readily be made by adjusting them in the light of a comparison of No. 8 and No. 14, which differ only in respect of emigration to the extent of an annual loss of 50,000 persons.

For the future the outlook is little clearer now than it was four years ago. The constant mortality basis may now be abandoned without danger, and an annual emigration of perhaps 50,000 may be envisaged, at least for a time; the numbers of births, however, may vary unpredictably and in regard to them the three projections chosen for mention in the main report still appear to be not inappropriate as possible alternatives.

REFERENCES

[1] 'Rise and Fall of the Perambulator Trade.' *Financial Times*, 8th September, 1951.

[2] Bogue, D. J. (1950). 'A Technique for making Extensive Population Estimates.' *J. Amer. Statist. Assoc.*, **45**, 149.

[3] Pearl, R. (1925). *The Biology of Population Growth.* Johns Hopkins University Press, Baltimore.

[4] Yule, G. U. (1924). 'The Growth of Population and the Factors which Control it.' *J. R. Statist. Soc.*, **88**, 1.

[5] Bowley, A. L. (1924). 'Births and Population in Great Britain.' *Econ. J.*, **34**, 188.

[6] Cox, P. R. (1950). *Demography* (Ch. 12 and 13). Cambridge University Press.

[7] *Idem* (1952). 'Population Projections: Notes on the Choice of Bases.' *J. Inst. Actu. Students Soc.*, **11**, Part 2.

[8] Lorimer, F. (1946). *The Population of the Soviet Union: History and Prospects.* League of Nations.

[9] *Report of the Royal Commission on Population* (1949). HMSO, London.

[10] *Papers of the Royal Commission on Population. Reports and Selected Papers of the Statistics Committee* (1950). HMSO, London.

A COLLABORATIVE INVESTIGATION INTO METHODS OF CHEMICAL ANALYSIS OF GROUND PYRETHRUM FLOWERS

GEORGE V. DYKE and EMILY P. SIMPSON (née Poulton)
Rothamsted Experimental Station, Harpenden, Herts

This article is of wide interest because it illustrates how a planned investigation with a statistical analysis of the results can disclose the sources of variations in the results of chemical analyses and give measures of the errors. Readers who are interested only in the practical conclusions may confine their attention to Part I; Part II is for those who are interested also in the statistical basis of the conclusions.

PART I. DESCRIPTION AND RESULTS OF INVESTIGATION

Introduction

Pyrethrum flowers are widely used in the preparation of insecticides. The active constituents of the flowers are a group of substances known as pyrethrins, and a bale of flowers is valued commercially according to the percentage content of all pyrethrins found in a sample of flowers taken from the bale. Over a period of years discrepancies were often noticed between the determinations of pyrethrin content made by the producers and those made by the purchasers of batches of flowers. It was known that these were largely due to the different methods of chemical analysis used by the parties concerned and to the variation to be found between samples taken from a batch; but it was also suspected that there were differences between the determinations when the same analytical process was used in different laboratories. The Colonial Products Advisory Bureau therefore decided to undertake an investigation of these differences, with the co-operation of producers and purchasers of pyrethrum flowers in fourteen different countries. A report issued by the Bureau[1] gives the results of this investigation and an account of the implications in insecticidal chemistry, and a summary of this has also been published.[2]

This investigation should be of interest not only to insecticide chemists but to all concerned with the standardisation of analytical techniques. The methods followed and the results achieved show what may be obtained from collaboration on an international scale, together with the difficulties encountered during the course of such an investigation. As temperature and humidity were believed to affect the pyrethrin determinations, laboratories were chosen in widely differing climatic conditions. The investigation also provided information on the relative magnitude of the variation between similar determinations in different laboratories, and that between duplicate determinations by the same

analyst in the same laboratory. In addition to this, some information was obtained on the reproducibility of a determination when made by analysts with greater or less experience.

Design of the Investigation

Two bales of pyrethrum flowers were selected, one of fresh flowers and one of flowers that had been stored for a year. Each bale was ground and carefully mixed to ensure uniformity, and one sample was drawn from each mixture and sent to each laboratory. Except in one case all the determinations at a laboratory were carried out by a single analyst. In general each analyst made twelve determinations in two sets of six. Each set consisted of one determination on each sample made by each of three chemical methods:

M (Mercury reduction), S (Seil), R (Ripert)

The arrangement within the set was random and was specified beforehand for each analyst. The interval between sets was determined by the individual analyst and varied from a few days to a month.

The amounts of pyrethrin I and pyrethrin II were determined separately. The total pyrethrin content is the sum of these two quantities. For various reasons some of the analysts did not carry out determinations by some of the methods. All results obtained have, however, been incorporated in the statistical analysis except where special mention is made.

Results

Means were computed from the determinations of pyrethrins I and II and total pyrethrins. The laboratories were grouped by regions according to the distance from the United Kingdom travelled by samples between grinding and analysis, and the climatic conditions to which the samples were exposed before they reached their destination. The grouping adopted was:

Group	*No. of Laboratories*
USA and Canada	15
UK	9
Europe (except UK)	6
Others*	12
Total	42

* The results grouped as 'others' came from laboratories in Africa, India, New Zealand, and S.E. Asia. In all these cases the samples were exposed to tropical conditions after grinding.

Table I shows the mean determinations of the four groups for total pyrethrin content. Corresponding results for pyrethrins I and II and the determinations of individual analysts may be found in the official report.[1] The three methods gave different mean determinations of pyrethrin content; method S gave consistently lower results than

method M, while method R gave lower results than method S for bale A but showed no consistent differences for bale B. Nevertheless for each method bale B was higher than bale A. The mean differences between bales (B — A) by the three methods were

M 0·08, S 0·11, R 0·18

These differences are influenced only to a small extent by the inclusion of determinations made by analysts who omitted altogether one or two of the methods M, S, and R. The statistical significance of the differences is discussed in Part II. It was found that method R gave on the average a significantly larger estimate of the difference than did the other two methods, while there is some suggestion of a smaller difference between methods M and S.

TABLE I

Mean values for determinations of total pyrethrin content

Method ..	M			S			R		
Bale	A	B	Mean	A	B	Mean	A	B	Mean
USA and Canada	1·35	1·40	1·38	1·13	1·23	1·18	1·06	1·26	1·16
UK	1·34	1·44	1·39	1·17	1.28	1·22	1·05	1·21	1·13
Europe (except UK)	1·33	1·46	1·39	1·10	1·22	1·15	0·98	1·22	1·10
Others	1·35	1·46	1·40	1·11	1·23	1·17	1·09	1·27	1·18
Mean	1·35	1·43	1·39	1·13	1·24	1·18	1·06	1·24	1·15

It was thought that tropical conditions might have some effect on the mean determination obtained from a laboratory and also on the variability both within and between analysts' results. From Table I it can be seen that the group named 'others' gave, on the average, slightly higher determinations by methods M and R than the remaining groups. In Part II it is shown, however, that these differences may be due entirely to chance effects.

Results of the Analysis of Variance

In order to assess the magnitude of the different sources of variation and to determine whether the differences already noted in the mean values have any real significance it is necessary to carry out analyses of variance of the individual results. Such analyses were made for pyrethrins I and II and total pyrethrin content, for each chemical method separately. The results of these analyses for total pyrethrins are presented in Part II of this article; we give here a summary of the conclusions that emerge.

The total variation in the determinations, apart from differences between bales, was found to be attributable to three main sources:

consistent differences between laboratories, variation between the estimates of the difference between the bales made by the various laboratories, and variation between duplicate determinations. For each method the variation between laboratories was much greater than would be expected from that within laboratories and, in fact, accounted for the major part of the total variation. The variation between laboratories was similar in the various geographical regions. The differences between bales also varied from laboratory to laboratory more than could be accounted for by variation within the laboratories.

Each analyst reproduced his results with a fair degree of precision, but the large variation between laboratories shows that far greater differences may be expected between two determinations made by different analysts in different laboratories than if they were made by one analyst in a single laboratory. (Since only a single analyst took part in the investigation in each laboratory it is impossible to say whether the differences that were observed between laboratories were primarily due to the analysts or to other factors differing from laboratory to laboratory.) The standard error for comparison between laboratories is thus considerably larger than that for comparison within laboratories; for while the latter arises from the residual variation within laboratories (i.e. variation between duplicate determinations) the former includes also the between-laboratory variation. Table II shows these two standard errors for each chemical method, together with the lower limit beyond which the standard error for comparison between laboratories cannot be reduced by increasing the number of determinations made by each analyst. For each method the standard errors are expressed as percentages of the mean determinations of total pyrethrins. The formulae from which these errors are derived are given in Part II.

TABLE II

Standard errors (per cent. of total pyrethrins)

Method	M	S	R
Within laboratories, single determinaton	2·8	2·8	5·0
Between laboratories, single determination	8·5	6·9	11·3
Between laboratories, mean of large number of determinations	8·1	6·3	10·2

The within-laboratory errors for methods M and S are both less than 3 per cent. of the mean determination, but that for method R is as large as 5 per cent. The between-laboratory errors are much greater, R being again the largest, and these errors are little reduced by increasing the number of determinations per laboratory.

Effect of Experience

An investigation was made into the effects of experience in the

chemical processes concerned, using only the results of those analysts who conformed strictly to methods M and S. Twenty-nine of the analysts conformed strictly to the standard M method and thirty-four to the standard S method. Analysts who had previously made ten analyses or had had two years' experience in the methods were classified as experienced and the rest as inexperienced. Seventeen analysts were classified as experienced in method M and twenty-three in method S. In some cases analysts who had considerable experience of one method had little or no experience of the other.

Means and standard errors were computed for the experienced and inexperienced in each method for pyrethrins I and II separately. The geographical regions were ignored in this analysis. There was no indication of consistent differences between the mean values of the determinations of the experienced and inexperienced analysts for either method. It was also found that although experienced analysts could repeat their results with a greater degree of precision than those with little experience of the methods, the variation between laboratories was greater for experienced than for inexperienced analysts. This last result is believed to be common to many types of chemical analysis, but in most cases the variation between the results of different analysts when each makes a large number of determinations is less than that between the results of different determinations by the same analyst. In the present case, however, the variation between analysts (or laboratories) is so much larger than that of separate determinations by the same analyst that the anomalous position is reached in which the standard error for comparison of two determinations made in different laboratories is larger for the experienced than for the inexperienced analysts. Table III shows these standard errors expressed as percentages of the mean determinations of all analysts conforming strictly to the standard methods.

TABLE III

Standard error (per cent.) of a single determination for comparison between laboratories for experienced and inexperienced analysts

Method	M		S	
Pyrethrins ..	I	II	I	II
Experienced ..	9·8	11·0	8·6	13·8
Inexperienced ..	8·4	10·6	7·0	8·6

PART II. STATISTICAL METHODS

Analysis of Variance

Analyses of variance were first carried out on the determinations of pyrethrins I and II by each of the methods of analysis for each regional group separately. The mean squares between laboratories were compared by the use of Bartlett's test for homogeneity of variances, and no

evidence was found to support the contention that the variation between laboratories differed substantially from group to group. A similar result was obtained from the comparison of mean squares within laboratories.

Analyses of variance of all the data were then carried out for each method and for pyrethrins I and II and total pyrethrins. Table IV shows the results of these analyses for total pyrethrins. In these analyses the results of one laboratory where the two replicates were obtained by different analysts were omitted for all methods; the results from two laboratories were also omitted from method M because of anomalous variation between replicates, possibly due in each case to a long interval between the analyses. If these two last results are included the variation within laboratories is considerably increased, doubling the mean square for laboratories × replicates (line 7) and considerably increasing that for laboratories × bales × replicates (line 9). Where an analyst carried out one replicate only by a particular method the result has been included in the normal way, allowing for a variable number of observations in a class.

TABLE IV

Analyses of variance for total pyrethrins. Degrees of freedom and mean squares (units of a single determination)

Source	Method M		Method S		Method R	
	D.F.	M.S.	D.F.	M.S.	D.F.	M.S.
1. Regional groups	3	0·00754	3	0·02912	3	0·03100
2. Laboratories within regions	35	0·05292	37	0·01967	30	0·05286
3. Laboratories	38	0·04933	40	0·02038	33	0·05087
4. Between bales A and B	1	0·32792	1	0·53993	1	1·17106
5. Laboratories × bales	38	0·00415	40	0·00413	33	0·01032
6. Replicates	1	0·00137	1	0·00113	1	0·01165
7. Laboratories × replicates	36	0·00115	40	0·00132	32	0·00287
8. Bales × replicates	1	0·00803	1	0·00002	1	0·00485
9. Laboratories × bales × replicates	36	0·00164	40	0·00087	32	0·00348
10. Between replicate determinations within laboratories	74	0·00148	82	0·00108	66	0·00333

Comparison of the mean squares of regional groups (line 1) with those of laboratories within regions (line 2) gives tests of significance for the regional means. There is no evidence that tropical conditions encountered after grinding affect the mean determinations of the analysts. In methods M and R the mean squares between regions are below expectation, and though this is not so for method S, the mean square between regions is considerably below significance level. We may in fact conclude that the variation between regional groups shown in Table IV is no more than can be accounted for by the variation between laboratories within regions. The degrees of freedom for regions and laboratories within regions were therefore pooled (line 3). It was also considered unnecessary to examine separately the interactions

involving regional groups, and these were accordingly amalgamated with the corresponding interactions within regions.

On comparing the mean squares for laboratories (line 3) with those due to variation within laboratories (lines 4–9) it becomes evident that differences between laboratories account for the major part of the variation other than that due to difference between the bales.

The mean squares for the difference between the bales (line 4) are clearly significant (tested by comparison with line 5), but the mean squares of line 5 (laboratories × bales) are also significantly large when compared with those of line 10. This shows that the estimates of the difference (B — A) between bales vary between laboratories more than can be accounted for by variation between the results obtained in the same laboratory.

Lines 6–9 are concerned with replicate differences and the variation of these differences between analysts and samples. The mean squares of line 6 measure the average difference between replicates for all the analysts by each method and are tested against those of line 7 which represent the variation in replicate difference between the analysts. The mean difference, if any, represents a time change, since replicate 1 was always analysed before replicate 2. In Table IV it can be seen that the mean squares between replicates exceed that for laboratories × replicates for method R only, and in this case the 5 per cent. significance level is not quite reached. The mean squares for laboratories × replicates (line 7) may be compared with the triple interaction laboratories × bales × replicates (line 9). It is evident from Table IV that for total pyrethrins the differences between these mean squares are small. Comparison of lines 8 and 9 tests the consistency between replicates of the differences between the samples. Method M gives evidence that the samples did not change similarly in the period between replicates, but methods S and R do not confirm this.

The mean squares of lines 6 and 9 for the three methods show no consistent differences, and it is therefore reasonable to pool them for each method to obtain estimates of the within-laboratory variances (line 10).

Standard Errors of Various Types of Comparison

The results of the foregoing analysis (Table IV) show that apart from the consistent differences between bales A and B the total variation of the results of each method can be attributed to three major sources: laboratory differences (line 4), variation in the estimated differences between bales A and B of different laboratories (line 6), and variations within a laboratory (line 10). Each determination may be expressed as the sum of the mean determination of the bale and three independent quantities which for the pth laboratory and qth bale (q = A or B) are :

(1) l_p. Constant for all the determinations of the pth laboratory. For the various laboratories the l_p ($p = 1, 2 \ldots 42$) can be regarded as a random variate distributed with zero mean and variance σ_l^2.

(2) i_{pq}. Constant for replicate determinations in the pth laboratory on bale A or B and distributed with zero mean and variance σ_i^2.

(3) ϵ. A random variate with mean zero and variance σ^2.

Then if L, I, and R represent the mean squares of lines 3, 5, and 10 respectively (Table IV) their expected values are:

$$4\sigma_l^2 + 2\sigma_i^2 + \sigma^2 \text{ and } 2\sigma_i^2 + \sigma^2 \text{ and } \sigma^2$$

Hence we estimate the variance components:

Variance	Estimate	Numerical Value of Estimate		
		Method M	Method S	Method R
σ^2	R	0·00148	0·00108	0·00333
σ_i^2	$\frac{1}{2}(I - R)$	0·00134	0·00152	0·00350
σ_l^2	$\frac{1}{4}(L - I)$	0·01130	0·00406	0·01012

The expected values of the mean squares given above are, however, subject to slight modification for methods M and R because of the omission of replicate determinations by two analysts for method M and by one for method R. The exact expectations are[3]:

$$L = \left\{\left(\mathcal{N} - \frac{p(4^2) + q(2^2)}{\mathcal{N}}\right) \Big/ (p + q - 1)\right\}(\sigma_l^2 + \tfrac{1}{2}\sigma_i^2) + \sigma^2$$

$$I = \left\{\left(\mathcal{N} - \frac{p(4^2) + q(2^2)}{\mathcal{N}}\right) \Big/ (p + q - 1)\right\}(\tfrac{1}{2}\sigma_i^2) + \sigma^2$$

where $\mathcal{N}$ = total number of determinations,

p = number of analysts carrying out four determinations in all by the given method, and

q = number of analysts carrying out two determinations in all by the given method.

The exact formulae lead to estimates of σ_l^2 and σ_i^2 for methods M and R slightly greater than those given above. The differences in the standard errors of Table II are negligible; all standard errors have been calculated from the unadjusted formulae.

The magnitude of the interaction between laboratories and bales may be expected to depend on the differences both between the laboratories and between the bales included in the tests—in general the greater the variation between the laboratories and between the bales the greater may we expect I to be. Although a reasonably representative sample of laboratories was included there were only two bales, and for this reason I, although based on 38, 40, and 33 degrees of freedom for methods M, S, and R respectively, cannot be regarded as well determined. Nevertheless in spite of the fact that the bales chosen differed widely in content, I is considerably smaller than L. Consequently we may reasonably suppose that I would probably not be

sufficiently large to affect greatly the estimates of error variance whatever bales might have been used.

It may be noted that although this investigation provides no estimate of the variance between successive determinations on the same bale by the same method the mean square R is unlikely to differ substantially from this variance.

If n determinations are made by one analyst in a laboratory chosen at random, the appropriate variance of the mean for comparison with similar means obtained in different laboratories is $\sigma_l^2 + \sigma_i^2 + \sigma^2/n$. However many determinations are made in each laboratory, this cannot be reduced below $\sigma_l^2 + \sigma_i^2$.

Three standard errors were calculated, these were:

(1) Standard error of a single determination for comparison within a laboratory: σ.

(2) Standard error of a single determination for comparison between laboratories: $\sqrt{(\sigma_l^2 + \sigma_i^2 + \sigma^2)}$.

(3) Lower limit of standard error of the mean of a large number of determinations by one analyst for comparison between laboratories: $\sqrt{(\sigma_l^2 + \sigma_i^2)}$.

These errors are shown in Table II.

Differences between Tests

It was pointed out in Part I that not only did the three methods give different mean values but they also gave different estimates of the difference between A and B. The errors of comparisons between methods are not covered by the analyses of variance already given, which deal with each method separately. Tests for such comparisons require the construction of analyses of variance based on the differences *between* determinations by the different methods. These can be obtained either by calculating the individual differences and analysing these or by the use of covariance. In the present case, however, there is no need to carry out full analyses of covariance, as the laboratory component of variation has been shown to be dominant.

An exact test of the variation in the difference (B — A) was made. As method R was omitted by a number of analysts the difference between methods M and S were first tested. For this purpose results were rejected which came from laboratories which did not carry out the complete set of determinations by both methods. The differences for each laboratory between the mean estimates of (B — A) by the two methods were listed and the mean of these differences tested by a *t*-test.

The mean difference so calculated is 0·026 with standard error (based on 36 degrees of freedom) 0·0135, giving a value for t of 1·93. This corresponds to a level of probability between 0·05 and 0·1. A similar comparison between the estimates of the differences by method R and the means of those by methods S and M, gives a mean of 0·072 with standard error 0·0194 (30 degrees of freedom), corresponding to a

probability of about 0·001. The conclusions drawn from these figures are given in Part I.

Effect of Experience

Part I describes the investigation into the effects of experience in the methods M, S, and R. In addition to the results presented there, Tables V and VI show the variation within and between laboratories for each class of analyst. They correspond to lines 10 and 3 in Table IV. As there is no *a priori* reason for assuming either class to have the lower variance a two-tailed test of significance should be used, and this implies doubling the probabilities obtained from the usual tables for z or F.

TABLE V

Mean squares ($\times 10^4$) *between replicate results of individual analysts, for those with and without experience in methods M and S*

Method	M			S		
Pyrethrin		I	II		I	II
	D.F.	Mean Square		D.F.	Mean Square	
Experienced	30	4·5	5·0	44	5·8	3·3
Inexperienced	20	14·4*	12·7*	22	9·1	6·5

* Significantly greater (at the 5 per cent. level) than corresponding mean square for experienced analysts.

TABLE VI

Mean squares ($\times 10^4$) *between laboratories for experienced and inexperienced analysts*

Method	M			S		
Pyrethrin		I	II		I	II
	D.F.	Mean Square		D.F.	Mean Square	
Experienced	14	212	176	21	101	171*
Inexperienced	10	137	79	10	58	54

* Significantly greater (at 5 per cent. level) than corresponding mean square for inexperienced analysts.

Table V shows that the mean squares for within-laboratory variation are larger for the inexperienced analysts than for those with experience of the methods, significantly so (at the 5 per cent. level) in two cases. Table VI shows that the mean squares between the determinations of experienced analysts are in each case greater than those between those of inexperienced analysts. The ratio of these mean squares is particularly large in the case of pyrethrin II by both the methods, reaching the 5 per cent. significance level for method S. Thus although on the

average the experienced analysts could reproduce their own results more closely than the analysts with little experience, the mean results of the experienced analysts varied more between themselves than those of analysts with little experience in these methods.

ACKNOWLEDGMENTS

The authors are indebted to the Colonial Products Advisory Bureau, and to the chairman and members of the Consultative Committee on Insecticide Materials of Vegetable Origin, for permission to publish results obtained in collaboration with them from the data provided by the analysts. They wish also to acknowledge the help and advice received from Dr. F. Yates at all stages of the investigation.

REFERENCES

[1] Colonial Products Advisory Bureau (Plant and Animal), Consultative Committee on Insecticide Materials of Vegetable Origin (1950). *Report of the Standing Sub-committee on Methods of Analysis of Vegetable Insecticides on the World-Wide Collaborative Analysis of Pyrethrum Flowers*, 1948–1949.

[2] Coomber, H. E., and Mitchell, W. (1951). 'The Active Principles of Pyrethrum Flowers.' *Analyst*, **76**, 556.

[3] Dyke, G. V. (1951). Appendix to Fenwick, D. W., and Franklin, M. T. 'Further Studies on the Identification of *Heterodera* Species by Larval Length. Estimation of the Length Parameters for Eight Species and Varieties.' *J. Helminth*, **25**, 57.

INDEX NUMBERS OF RETAIL PRICES 1938–51

ROY G. D. ALLEN
London School of Economics

This article is based largely on the work of the Cost of Living Technical Committee as summarised in the recently published *Report on the Working of the Interim Index of Retail Prices.* Here index numbers are calculated with alternative weights. Links are worked out for extending back to June 1947 and to 1938 the new series of index numbers with January 1952 as 100, as published by the Ministry of Labour. Professor Allen is a member of the Technical Committee.

Introduction

The Cost of Living Advisory Committee has reported the findings of its Technical Committee on the operation of the Interim Index of Retail Prices for working-class and lower-salaried families.[1] The index, as calculated by the Ministry of Labour until January 1952, was weighted according to a pre-war (1937–8) pattern of consumption but showed price changes only from June 1947. The conclusions of the Committee were:

(*a*) that substitution of a post-war pattern of consumption as weights would reduce slightly the index figure, to show a rather smaller percentage change in all retail prices since June 1947, and

(*b*) that the index ignored the higher rents for some of the houses newly let after June 1947 and that it may not have been able, completely or in all cases, to record price increases arising from quality depreciation, i.e. it understated slightly the change in all retail prices since June 1947.

These two factors broadly offset each other in the all-items index, and the published figure ($132\frac{1}{2}$) for January 1952 is a fair representation of the price increase from June 1947. However, this may not remain true of the all-items index in future and it is certainly not true of the index numbers for all the separate groups of expenditure which make up the total. Consequently the Committee recommended that a revised index be introduced with a post-war weighting pattern based on 1950 consumption, with certain modifications in the price series for rents and some other items, and with a new base date. These recommendations have been adopted, and the index now published is expressed on January 1952 as 100.[2] The new all-items index can be carried back to June 1947 by multiplying by the previous index ($132\frac{1}{2}$) for January 1952 on June 1947 as 100. No such linking can be done for the separate group index numbers which make up the all-items index.

The main purpose of the present paper is twofold. In the first place, alternative sets of weights, based on consumption patterns for working-

class and lower-salaried families of 1937–8, 1947, and 1950, are available in the Committee's report.[3] The effect of the use of these is examined by calculating index numbers from 1938 as well as from June 1947. Secondly, as a result of this examination an attempt is made to provide links for the current series of index numbers (for working-class and lower-salaried families) with January 1952 as 100, to take them back to June 1947 and to 1938. The links are given, not only for the all-items index, but also for all the individual groups of expenditure distinguished in the index.

The Basic Data

These are given in the appendix to this paper. Each of the groups and sections of the price index is shown separately, except that a few minor amalgamations of sections are made where estimates of price changes from 1938 to 1947 cannot be given in sufficient detail. Of the three consumption patterns given, that of 1937–8 is based on the results of an enquiry into family budgets of wage-earners and small-salary earners. The patterns of 1947 and 1950 are obtained by adjusting that of 1937–8 for changes in national consumption between 1938 and 1947 or 1950, as shown in the White Papers on National Income and Expenditure, and by using other information available on working-class expenditure in the post-war period.[4] The three patterns of consumption must be shown at the same set of prices; this is done in the Appendix in alternative ways by pricing all three patterns first at June 1947 prices and then at August 1951 prices.

Price changes for each section of expenditure are given for two periods, average 1938 to June 1947 and June 1947 to August 1951, and hence by multiplication for the combined period from average 1938 to August 1951. The price relatives for the earlier period covering the war years are estimates which are necessarily rough. They take into account price indices from the White Papers on National Income and Expenditure and from the old cost-of-living index, together with what can be deduced from the weights of the Interim Index of Retail Prices as analysed by Dudley Seers.[5] The price relatives for the period since June 1947 are the sectional index numbers of the Interim Index of Retail Prices.

Two particular price relatives require explanation. The first is that for rent and rates, where the procedure recommended by the Advisory Committee and adopted in the revised index is followed here throughout. Consequently the price relative for August 1951 (June 1947 = 100) is taken as 108½, i.e. the group index of 104 from the Interim Index on the basis of rents of houses in existence in June 1947, raised by 'between four and five points' to allow for rents of dwellings subsequently built for letting.[6] The second is the price relative for beer, where the practice of the revised index, substantially unchanged from the previous index, is followed. The unit of quantity adopted for beer can be either the 'bulk barrel' or the 'standard barrel'; in the former case the price

makes no allowance and in the latter case it makes full allowance for changing strength of beer. For years up to 1946 the White Papers on National Income and Expenditure showed the results of using both units; the price relative for beer in 1946 (1938 = 100) was 209 on the 'bulk barrel' basis and 258 in terms of 'standard barrels.'[7] The compromise adopted here, as in the revised index, is to take a straight average of the two price relatives (e.g. $233\frac{1}{2}$ in 1946).

The assumption throughout this paper is that the price relatives for sections of expenditure can be treated as price quotations proper, independent of any system of weighting within the sections. This is a reasonable assumption, since weighting in the retail price index is generally taken only to the section level; within each section several price quotations are selected and combined, either without weighting or with only the roughest form of weighting. There are, however, two or three sections where internal weighting has some effect on the price index for the section, e.g. fruit and vegetables (taken as one combined section in the basic data here) and travel. These effects are, in fact, ignored in the present calculations; they make very little difference to the main group and all-items index numbers.

Consumption Patterns

The three consumption patterns for working-class and lower-salaried families are shown, in detail and in the groupings of the retail price index, in the appendix table. They are summarised in Table I in a rather different arrangement, roughly in conformity with the groupings of the White Papers on National Income and Expenditure. For a given pattern of expenditure, which is a quantitative concept, a percentage distribution over different items can only be obtained if some fixed set of prices is used to aggregate the quantities. The percentage one item bears to the total will vary according to the particular set of prices used. But changes in the relative importance of an item from one consumption pattern to another are adequately disclosed whatever the set of prices, provided only that the same set is used in valuing up each consumption pattern. Hence, to see changes in relative importance, Table I needs to be read across the rows for either set of valuations (at 1947 or at 1951 prices).

The greatest interest lies in a comparison of the 1950 pattern of consumption with that of 1937–8. The comparison confirms broad expectations. The relative consumption of rationed foods, coal, furniture, soft furnishings and clothing (excluding footwear) was much lower in 1950 than before the war, while drink and tobacco consumption was also a rather smaller proportion of the total. Consumption increased particularly for non-rationed foods, gas and electricity, newspapers and books, and services. The pattern of consumption in 1947, on the other hand, was peculiar in many ways, differing both from the pre-war and from the 1950 pattern. The differences, moreover, were associated with price changes in a manner far from normal. For

TABLE I

Patterns of consumption, working-class and lower-salaried families, 1937–8, 1947, and 1950

	Interim Index Groups and Sections	Consumption at June 1947 prices			Consumption at August 1951 prices			Price Relatives†		
		1937–8	1947	1950	1937–8	1947	1950	June 1947 (Average 1938 = 100)	August 1951 (June 1947 = 100)	August 1951 (Average 1938 = 100)
Rationed food*	I	131	82	96	167	106	125	114½	163	187
Other food	I	217	277	264	215	279	263	157	126	197
Alcoholic drink	VIII	101	101	93	86	88½	80	230	109	251
Tobacco	IX	116	124	111	95	104½	92	360	105	377
Rent and rates	II	88	87	87	75	76	75	110	108½	119
Coal and coke	IV 1	41	28	27	41	29	27	160	127	203
Other fuel and light	IV 2, 3, 4	24	35	37	24	35	37	126	129	162
Miscellaneous household goods, etc.	VI 1, 2, 3	24	20	23	23	20	23	153	125	191
Furniture, hardware, etc.	V 1, 2, 5	51	39	42½	49	38	41	206	122	250
Soft furnishings	V 3, 4	20	13	17½	28	19	25	238	178	424
Clothing (excluding footwear)	III 1–7	77	56	64	87	65	74	175	144	253
Footwear	III 8, 9, 10	20	20	23	22	22	25	174	139	242
Newspapers, books, etc.	VI 4	11	16	18	13	19	21	120	147	176
Travel	VII 1	25	32	31	24	32	30	130	123	160
Entertainment	VII 3	18	27	26	17	25	24	175	115	201
Other services	VII 2, 4	36	43	40	34	42	38	156	121	188
		1,000	1,000	1,000	1,000	1,000	1,000			

* Foods at present on ration: Beef, mutton and lamb, bacon and ham, butter, margarine and cooking fat, cheese, eggs, tea, sugar.

† Weighted (where necessary) with 1937–8 weights.

example, the consumption of rationed foods was very low in 1947 at a time when prices were also relatively low; price rises occurred mainly after 1947, when consumption was allowed to increase. For this reason 1947 consumption cannot provide a good system of weighting for a retail price index; it would, for example, understate the effect of price increases among rationed foods in the period after 1947.

When the actual proportion of one item of expenditure in the total is measured for a given consumption pattern the particular set of prices used is a matter of great importance. So in the 1950 budget rationed food represents 12½ per cent. and clothing and soft furnishings 10 per cent. of the total when 1951 prices are used; at 1947 prices the proportions are only 10 per cent. and 8 per cent. respectively. This is because price movements from 1947 to 1951 were diverse, with rationed food, clothing, and soft furnishings all increasing in price far more than the average. This is a point which is often misunderstood, without the slightest excuse. It is not correct to say that there was little relative change in the importance of rationed foods from pre-war to 1950 because these items represented 12½ per cent. of the 1950 budget (at 1951 prices) and only 13 per cent. of the pre-war budget (at 1947 prices as in the weights of the Interim Index of Retail Prices). The prices used are different and affect the comparison. The same point arises when an index (on a given pattern of consumption) is calculated first with one and then with another period as 100. Suppose that the index is computed on the basis of the 1950 pattern of consumption. Then, if June 1947 is taken as 100, the weight of rationed food is 9·6 per cent. (i.e. 1950 consumption at June 1947 prices), but if August 1951 is 100 the weight is 12·5 per cent. (i.e. 1950 consumption at August 1951 prices). This is not a change in weighting, for the index is fixed on the basis of 1950 consumption; it is merely the effect of a different starting point (with different prices) used in a given index.

Price Index Numbers for Working-class Families

Table II shows index numbers of price changes from average 1938 to June 1947 and to August 1951 for all retail prices and for each of the main groups. Each index is computed on the basis of the three different patterns of consumption.

The first conclusion is a confirmation that the 1947 budget is not a suitable basis for weighting the all-items index either in the period before 1947 or in that after 1947. It overstates the general price rise from 1938 to 1947 and understates the subsequent increase in prices. The effective choice is between the pre-war (1937–8) and the 1950 weightings. The differences between these two bases in the all-items index are not great,* and what differences exist are in the usual and

* There are only a few substantial differences even among the group index numbers. In the period 1938–47 the 1950 weighting gives a higher index for food because of the greater weight attached to non-rationed foods with large price increases, and a lower index for fuel and light because of greater weight given to electricity (which fell in price).

TABLE II

Price index numbers, working-class and lower-salaried families, different weighting systems

	June 1947, Average 1938 = 100			August 1951, June 1947 = 100			August 1951, Average 1938 = 100		
	Consumption pattern of:			Consumption pattern of:			Consumption pattern of:		
	1937–8	1947	1950	1937–8	1947	1950	1937–8	1947	1950
Rationed food	114½	116	115	163	160	161	187	185	185
Other food	157	158	158	126	125	126	197	198	198
I. Total food	138	146	143	140	133	135	192	194	194
II. Rent and rates	110	110	110	108½	108½	108½	119	119	119
III. Clothing	175	175	175	143	142	143	250	249	250
IV. Fuel and light	145½	134	132	128	126½	126	186	169	166
V. Household durable goods	214	212	214½	138	135	139	294	285	297½
VI. Miscellaneous goods	141	136½	136	132	135	135	186	184	184
VII. Services	150	151	151	120	120	119	181	181	180
VIII. Alcoholic drink	230	230	230	109	109	109	251	251	251
IX. Tobacco	360	360	360	105	105	105	377	377	377
All items	161·4	163·4	160·8	127·4	124·0	125·9	205·7	202·6	202·3

expected direction. The base-weighted (Laspeyres) index numbers on the 1937–8 pattern of consumption are rather greater than the current-weighted (Paasche) index numbers with the later (1950) pattern.

It follows that either set of weights (1937–8 or 1950 consumption) can be used reasonably well for measuring movements of general retail prices in the period from 1938 to 1947 and to 1951. A 'cross' between the two sets of weights would be appropriate, e.g. the Fisher 'ideal' form. The official index in its revised form from January 1952 onwards is based on the 1950 pattern of consumption. The question is, how to link this back to earlier dates. As a link for the relatively short period back to June 1947 it would appear to be quite appropriate to use the 1950 pattern of consumption and hence to obtain an index with the same weighting system throughout. For the longer link back to 1938, where the price comparison must necessarily be more approximate, the 'cross' between 1937–8 and 1950 weights has much to recommend it.

Links for the Revised Index at January 1952

The latest date of the index numbers of Table II is August 1951; they need to be brought up to January 1952 to provide links with the revised index calculated with January 1952 as 100. This can be done by applying to each section index at August 1951 the percentage change in prices shown by the Interim Index of Retail Prices between August 1951 and January 1952.* The results are set out in Table III. One decimal place has been carried throughout the calculations and shown in the final results, not because it has any significance in interpreting the index numbers but simply to facilitate any further computations which may be necessary, e.g. in linking with the revised index.

TABLE III

Price index numbers, working-class and lower-salaried families, January 1952

	June 1947 = 100 1950 pattern	Average 1938 = 100		
		1937–8 pattern	1950 pattern	Fisher 'ideal'
I. Food	143·9	206·2	206·4	206·3
II. Rent and rates	108·8	119·7	119·7	119·7
III. Clothing	146·2	257·6	255·9	256·7
IV. Fuel and light	136·5	204·0	179·6	191·4
V. Household durable goods ..	137·9	292·0	295·8	293·9
VI. Miscellaneous goods ..	141·8	193·2	193·2	193·2
VII. Services	123·1	186·2	185·8	186·0
VIII. Alcoholic drink	110·2	253·4	253·4	253·4
IX. Tobacco	107·1	385·6	385·6	385·6
All items	131·0	214·5	210·6	212·5

* The Ministry of Labour *Gazette* of March 1952 gives the 1950 weights valued at prices of January 1952. The reciprocals of the index figures in the first column of Tables III and IV are to be combined with these weights.

The index numbers of Table III take account of the adjustment in the series for rent and rates, and so include rents of houses newly let since June 1947 (as in the revised index). They make no adjustment for hidden price increases through any depreciation in quality of goods and services which the Interim Index of Retail Prices was not able to incorporate. In the view of the Advisory Committee, the published index up to January 1952 may understate 'to a slight extent' the true rise in prices on this account.[8] Some adjustment of the index numbers of Table III is needed. An estimate of the extent of the adjustment in any group index would require information which is not available; indeed, there are no guides beyond general impressions. Consequently nothing more than rather arbitrary adjustments are attempted here, with the results shown in Table IV.

TABLE IV

Price index numbers, working-class and lower-salaried families, January 1952. *Adjusted for quality depreciation, June* 1947 *to January* 1952

	June 1947 = 100 1950 pattern	Average 1938 = 100		
		1937–8 pattern	1950 pattern	Fisher 'ideal'
I. Food	146·1	209·6	209·5	209·5
II. Rent and rates	108·8	119·7	119·7	119·7
III. Clothing	150·8	265·8	263·9	264·8
IV. Fuel and light	138·3	207·0	182·0	194·1
V. Household durable goods ..	139·8	295·8	299·9	297·8
VI. Miscellaneous goods ..	143·9	195·8	196·1	195·9
VII. Services	124·3	188·0	187·6	187·8
VIII. Alcoholic drink	110·2	253·4	253·4	253·4
IX. Tobacco	107·1	385·6	385·6	385·6
All items	132·5	217·3	212·9	215·1

In the period June 1947 to January 1952 it is assumed that there is no understatement of the price rise for rent, drink, and tobacco, that the percentage *increase* in clothing prices is understated by 10 per cent., and that the percentage *increases* in other groups are understated by 5 per cent. Thus the recorded increase in food prices is 43·9 per cent., and this is raised by 5 per cent. to 46·1 per cent. No adjustments are made for any unrecorded quality depreciation in the period covering the war years from 1938 to June 1947. The approximate nature of any price comparison over this period is fully stressed in the White Papers on National Income and Expenditure and the same qualifications apply to all index numbers shown here with 1938 as 100.

The first column of Table IV provides links which are appropriate in carrying the revised index (January 1952 = 100) back to June 1947. The result is an index based throughout on the 1950 pattern of consumption and with adjustments (though rather arbitrary ones) for quality depreciation between June 1947 and January 1952.

The last column of Table IV gives what are perhaps the best links for carrying the revised index back to 1938. However, whatever is done, it must be remembered that the comparison between 1938 and the present can only be of a very rough and ready nature. Little allowance can be made for differences in the types and qualities of goods and services available in the pre-war and post-war periods. The links of Table IV are given to one decimal place for purposes of calculation; they are not likely, in fact, to be correct even to the nearest whole number.

Price Index Numbers for Middle-class Families

The White Papers on National Income and Expenditure show index numbers of retail prices computed for all consumers. These can be related to the index numbers (for working-class and lower-salaried families) described above, so that, by difference, index numbers for all other, or 'middle-class,' families can be obtained. The two series of index numbers must be made comparable both as regards their weighting systems and in respect of the treatment of beer prices. Such a comparison has been made by Mr. Dudley Seers[9] in terms of index numbers with 1938 weighting and with beer prices on a 'standard barrel' basis; in 1948, on this basis, the all-consumers index was 186½, the working-class index 175½ and the middle-class index 199 (1938 = 100). If the compromise solution is adopted for beer prices, as in the index numbers calculated above, these figures reduce to 185, 173, and 199.

A similar comparison for the period 1948 to 1950–1 can be made best in terms of current-weighted index numbers. The White Paper series for all consumers now has current year weights with 1948 as 100. A comparable working-class index can be written with 1950 weights and switched from June 1947 to 1948 as 100.* There is no adjustment for quality depreciation in the index numbers. The comparison is set out in Table V and summarised in Table VI.

TABLE V

	1948	1950	1951
All-consumers index†	100	105·9	114·1
Working-class index‡			
With June 1947 = 100	107·1	113·1	123·5
With 1948 = 100	100	105·6	115·3
Middle-class index§	100	106·3	112·3

* The all-consumers index now adopts the compromise treatment of beer prices, as in the working-class index. (See *Monthly Digest of Statistics*, March 1952.)

† From *Monthly Digest of Statistics*, March 1952.

‡ Computed in the first column of Table III.

§ By difference, estimating middle-class and working-class consumption expenditures in the ratio 2 : 3.

TABLE VI

Price index for:	1938 Weights		1950–51 Weights		
	1938	1948	1948	1950	1951
All consumers	100	185	100	106	114
Working class	100	173	100	106	115
Middle class	100	199	100	106	112

By 1948 there was a large differential between price rises for middle-class and working-class families; this has been reduced slightly after 1950 during a period of rising prices of subsidised foods and utility items.

More accurate index numbers for middle-class families, as for other special groups such as old-age pensioners, can only be calculated on the basis of fuller information on consumption patterns and on prices paid by various types of consumers. The results of the family-budget enquiry shortly to be undertaken will supply some of the information. But the need can only be fully satisfied if there are continuing budget enquiries and wider collections of price data covering different groups of consumers.

REFERENCES

[1] *Report on the Working of the Interim Index of Retail Prices.* Cmd. 8481, HMSO, 1952.
[2] Ministry of Labour *Gazette,* March 1952.
[3] Cmd. 8481, Appendices II and III.
[4] *Ibid.*, p. 7.
[5] SEERS, DUDLEY (1949). *Changes in the Cost of Living and the Distribution of Income since* 1938 (Blackwell, Oxford), Table III, pp 36–8.
[6] Cmd. 8481, p. 18.
[7] Cmd. 7099, p. 47.
[8] Cmd. 8481, pp. 20–9.
[9] SEERS, DUDLEY (June 1950). Article in *Bull. Oxford Inst. Statistics.*

APPENDIX

Consumption patterns and price relatives, working-class and lower-salaried families

Group and Section	Consumption at June 1947 prices			Consumption at August 1951 prices			Price Relatives		
	1937–8 (1)	1947 (2)	1950 (3)	1937–8 (4)	1947 (5)	1950 (6)	June 1947 (Average 1938 = 100) (7)	August 1951 (June 1947 = 100) (8)	August 1951 (Average 1938 = 100) (9)
I. Food									
Beef	25	18	18	32	23	23	120	161	193
Mutton and lamb	12	11	11	16	15	15	110	166	183
Bacon and ham	18	6	12	21	7	14	120	152	182
Butter	19	7	9	28	11	14	100	188	188
Margarine, cooking fats	12	8	10	14	9	11	160	144	230
Cheese	5	5	5	5	6	6	92	140	129
Eggs	13	6	11	20	10	18	92	201	185
Tea	16	13	13	15	13	13	125	122	152½
Sugar and syrup	11	8	7	16	12	11	125	186	232½
All rationed	131	82	96	167	106	125			
Bread and flour	30	33	30	32	35	32	110	134	147
Biscuits, cakes, etc.	32	43	36	29	40	33	275	115	316
Other meat	23	23	27	28	29	33	145	155	225
Fish	8	19	13	11	26	17	115	169	194
Fresh milk	32	35	39	32	36	39	130	126	164
Vegetables, fruit	52	71	64	44	60	55	175	108½	190
Other foods	40	53	55	39	53	54	180	124	223
All non-rationed	217	277	264	215	279	263			
All food	348	359	360	382	385	388			
II. Rent and Rates	88	87	87	75	76	75	110	108½	119
III. Clothing									
Men's	30	20	26	36	25	31	176	152	268
Women's	31	24	25	31	25	26	175	129	226
Children's materials	16	12	13	20	15	17	175	158	276½
Footwear	20	20	23	22	22	25	174	139	242
All clothing	97	76	87	109	87	99			

IV. Fuel and Light									
Coal and coke	41	28	27	41	29	27	160	127	203
Gas	14	19	18	14	20	19	145	131	190
Electricity	6	13	15	5½	12	14	90	116	104
Other fuel and light	4	3	4	4½	3	4	150	139	208½
All fuel and light	65	63	64	65	64	64			
V. Household Durable Goods									
Furniture	18	11	16	19	12	17	220	134	295
Appliances, etc.	24	20	19½	22	19	18	200	117	234
Floor coverings	11	7	9½	13	8	11	260	146	380
Drapery, etc.	9	6	8	15	11	14	215	218	469
Hardware, etc.	9	8	7	8	7	6	195	109	213
All household durable goods ..	71	52	60	77	57	66			
VI. Miscellaneous Goods									
Soap	8	6	8	8	6	8	140	132	185
Other goods	16	14	15	15	14	15	160	122	195
Newspapers, books, etc.	11	16	18	13	19	21	120	147	176
All miscellaneous goods	35	36	41	36	39	44			
VII. Services									
Travel	25	32	31	24	32	30	130	123	160
Postage, etc.	9	12	12½	7	10	10	160	100	160
Entertainment	18	27	26	17	25	24	175	115	201
Other services	27	31	27½	27	32	28	155	128	198
All services	79	102	97	75	99	92			
VIII. Alcoholic Drink	101	101	93	86	88½	80	230	109	251
IX. Tobacco	116	124	111	95	104½	92	360	105	377
All items	1,000	1,000	1,000	1,000	1,000	1,000			

Cols. (1), (2), (8). Cmd. 8481, Appendix II.
Col. (3). (6) divided by (8) scaled to total 1,000.
Col. (4). (1) times (8) scaled to total 1,000.
Col. (5). (2) times (8) scaled to total 1,000.

Col. (6). Cmd. 8481, Appendix III.
Col. (7). From available data. *Cf.* Dudley Seers.[5]
Col. (9). (7) times (8).

CALIBRATION OF AN INTEGRATING GAS FLOWMETER

WILLIAM SPENDLEY
Imperial Chemical Industries Ltd., Billingham

This article is an interesting fellow to Mr. Jessop's article *One Line or Two?* since it illustrates by a practical example the problem of choosing the most suitable line in regression analysis. It also illustrates the treatment of heterogeneous data. Mr. Spendley describes how corresponding readings of gas flow measured by two types of flowmeter were used to obtain a calibration curve for one of the types and to estimate the errors involved in using this curve.

Introduction

The first stage in the synthesis of ammonia by the Haber process is the production from coke and steam of 'water-gas' consisting mainly of hydrogen, carbon monoxide, nitrogen, and carbon dioxide. Blasts of air and steam are passed alternately through a bed of hot coke enclosed in a water-gas generator, the air to maintain—by combustion—the temperature of the coke bed, and the steam to produce the desired water-gas. In order that factors affecting the efficiency of water-gas generators might be investigated, one of the twenty generators at the Billingham works of ICI Ltd. was set aside for experimental purposes, and preliminary studies were begun both of the means by which generator efficiency might best be measured, and of the precision of measurement possible. An important problem which arose was that of measuring the volume of gas made. Since the experimental generator was still 'on line,' i.e. producing gas for synthesis purposes, its output was led to the sulphur-removal plant and raw gas holder in common with that from all other generators and could not be measured separately. Indirect determination was possible since there was an orifice-plate flowmeter in the main leading from the generator, but there were overriding practical difficulties in the use of this instrument. The flow of gas from a generator varies between wide limits; it is zero during the 'blow' (air blast) period, rises to a peak at the beginning of the 'make' period, fluctuates sharply thereafter, and drops to zero again when the next 'blow' period commences. When using the orifice-plate flowmeter at least 40–50 readings of the instantaneous rate of flow are necessary per cycle of 3–4 minutes (i.e. readings must be taken at least every 5 seconds), and numerical integration of the figures is extremely tedious. For runs of more than a few cycles' duration the work involved therefore becomes prohibitive. Measurements were, however, required of

the volume of gas made per day, per week, etc., so that some alternative method had to be devised. After consideration of various schemes it was decided that the best solution was to tap off through a smaller-diameter pipe a portion of the gas flowing in the main and to mount in this shunt an integrating flowmeter of the type (similar to a domestic gas-meter) in which an anemometer is driven by the gas stream and its revolutions counted. There remained, however, the problem of calibrating the instrument; the precise fraction of the total gas diverted through it was unknown, while its response to pulsating flow was unlikely to accord with that for steady flow. It is with the calibration of this integrating flowmeter that we are here concerned.

Experimental Procedure

To provide data for the calibration a series of runs was made in which the output from the generator for each of ten successive cycles was determined explicitly (using the orifice-plate flowmeter in the main) and was compared with the corresponding increase in the cumulative reading of the integrating flowmeter. Each run lasted some 30–40 minutes and involved the taking of some 400–500 separate readings. Longer runs would have placed undue strain on the operator, but the data for a single run were never adequate for calibration purposes since it was inconvenient to vary the generator output per cycle much over the relatively short period concerned. In all, therefore, twenty daily calibration runs of ten cycles each were made; these covered a sufficiently wide range of outputs for a calibration to be attempted.

When the results obtained from the calibration runs were plotted they displayed (as was to be expected) strong evidence of a linear relationship between shunt flowmeter reading and volume of gas as determined from the orifice-plate figures. There was, however, a fairly wide scatter of the experimental points about any line drawn through them, and statistical methods had therefore to be used to determine the 'best' calibration curve (i.e. that most truly representative of the experimental results) for the estimation of gas volume from shunt flowmeter reading.

It was decided to use standard regression methods to determine a linear calibration curve, but the problem arose, which regression line should be used ? This kind of problem arises so often that it is worth considering in some detail the arguments for and against each of the possible representations of the data of the present example, and the reasons which led to the final choice. The arguments are not invented; the two members of the Statistical Section at Billingham did not at first see eye to eye on the problem, and it was only after some discussion that agreement was reached.

Which Regression ?

Calibration problems have one characteristic not shared by other regression problems—the two variables concerned are not distinct but

are merely different versions of the same thing. This is generally true, although in the present example it tends to be obscured by the fact that the whole volume of gas is determined by the orifice-plate method, while the shunt flowmeter only measures a portion of the whole. There is, therefore, no obvious causality or dependence of one variable on the other. In the present example, additionally, there were known to be errors (mechanical and observational) in both sets of figures; the magnitudes of these could be estimated from results for those runs where no significant correlation between orifice-plate and shunt flowmeter figures were found, i.e. those runs in which there seemed to have been no important variations in output per cycle. Thus neither regression could be discarded *a priori*, following the usual textbook rules. One suggestion which was considered was that neither regression was in fact valid; since there were errors of measurement in both variables both lines gave biased estimates of the true relationship. This, however, was soon rejected; the true relationship was a measure of the way in which the gas stream divided at the junction, but was of no use for calibration purposes unless 'true' values of shunt flowmeter reading were available —a somewhat Utopian requirement ! The choice therefore lay between the regression of shunt flowmeter reading on volume of gas measured by the orifice-plate, and that of orifice-plate result on shunt flowmeter reading. In support of the former was the fact that the orifice-plate result was more closely related to actual gas flow, and the fact that it was a consistent measure from run to run; occasions might arise in which replacement of parts of the shunt flowmeter became necessary, causing a change in its response, while the data available showed that minor variations in the calibration from day to day occurred even without such interference. In support of the latter was the fact that prediction was required of the gas volume (i.e. 'true' orifice-plate result) from an observed shunt flowmeter reading, and the fact that the shunt flowmeter readings were subject to rather smaller proportionate errors than the orifice-plate figures. Discussion centred round these points for some time, and the further argument was adduced that the estimated change in gas volume for a given change in shunt flowmeter reading was less using the regression of orifice-plate result on shunt flowmeter reading than using the other regression. Though theoretically irrelevant, this was practically important since it was better to err on the side of caution in experiments designed to detect changes in efficiency. Eventually, however, a decision was reached without recourse to this argument. It was concluded that, for prediction from an *observed* value of a given variable, the regression of the other variable on that variable was appropriate whether or not it was an unbiased estimate of a functional relationship, and that, so long as observed values only were considered, the functional relationship was irrelevant—its estimation was a separate problem which should not be confused with the fitting of regression lines. Hence the regression of orifice-plate result on shunt-flowmeter reading should be used.

Statistical Analysis

Having decided how the data should be treated there remained the problem of analysing them statistically. The analysis, though not the subject of controversy, is nevertheless of some interest since a rather full treatment was necessary before satisfactory estimates of the error of prediction could be obtained. As mentioned above, 20 separate sets of 10 cycles were studied and each could in theory be used to give an estimate of the regression line. In 15 of the 20 sets, however, no significant regression was found, i.e. the shunt flowmeter and orifice-plate results for these runs were uncorrelated. Clearly it was not often that the output per cycle varied appreciably over the period of a single run, and the data for all twenty runs would therefore have to be combined to give a satisfactory estimate of the regression line. There were, however, some doubts as to the homogeneity of the data; the gas measured was very dirty, and it was thought possible (and was in fact confirmed by the analysis) that the response of the shunt flowmeter might vary from day to day because of random changes in the effects of inertia and friction. For this reason the data were analysed fully; the variation of the orifice-plate results was split up into that which could be associated with changes in shunt flowmeter reading (i.e. that accounted for by the overall regression), that associated with changes in the slope of the regression line between runs, that associated with changes in the position of the regression line between runs, and that attributable only to errors of measurement and observation. The analysis of variance table obtained is given in Table I.

TABLE I

Analysis of variance of calibration data

Source of Variation	D.F.	Sum of Squares	Mean Square
Overall regression	1	75,304·93	75,304·93
Difference in slope between regression of means and 'average within' regression	1	314·39	314·39
Variation of means about regression of means	18	14,614·14	811·90
Differences in slope between regressions within runs	19	1,239·24	65·22
Residual (i.e. error)	160	7,679·08	47·99
Total	199	99,151·78	—

The residual sum of squares is the sum of squares of deviations from the separate regression lines fitted to the results for each run. The mean square representing differences in slope between the regressions for each run is not significantly greater than the residual mean square, i.e. the observed differences in slope are no greater than could have arisen by

chance as a result of random error. There is therefore no evidence of real changes in slope between runs.

The second and third mean squares down the table reflect differences in the position of the regression line between runs; both, but particularly that representing the variation of means about the regression of means,

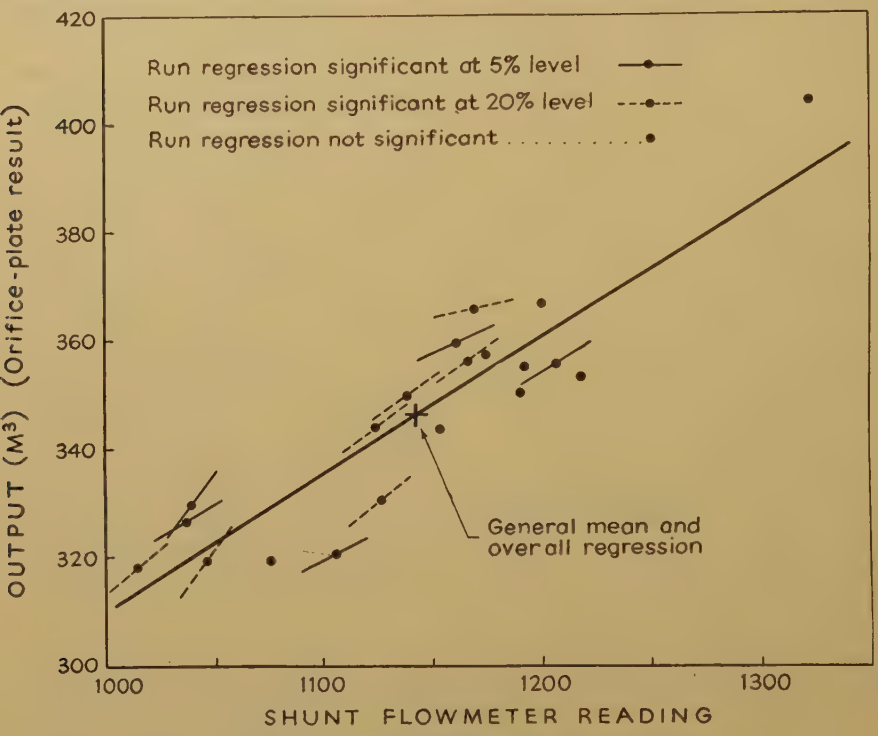


Fig. 1. Means and slopes of regressions for individual runs.

are highly significant compared with the residual mean square. This, though not entirely unexpected, is disturbing; it implies that real changes in position between runs must be assumed and that the data cannot be considered homogeneous. In the circumstances the heterogeneity must be accepted and due allowance must be made for it; it cannot—as would usually be the case— be made an excuse for rejecting the data. It is fortunate, therefore, that the variation of means about the regression of means is sufficient for the apparently significant difference between the slopes of the regression of means and the 'average within' regression to be ignored. As is confirmed from Fig. 1, there is no reason to suspect non-linearity, and the use of the overall regression line as a calibration curve is still justified. The errors involved are, however, discussed below; they are very different from those which would obtain if the data were homogeneous.

Prediction

Though there is heterogeneity of the data this is fortunately of a kind

which can be handled fairly easily. Two sources of variation only must be considered— the random errors of measurement, and the additional day-to-day variation of the run means. The variance associated with the former is the error variance of 47·99 found earlier. The expectation of the mean square representing variations of means about the regression of means is:

$$\sigma_o^2 + 10\sigma_m^2$$

where σ_o^2 is the error variance, σ_m^2 is the variance of the additional variation of run means, and the factor 10 arises because ten cycles were observed in each run. Hence an estimate of σ_m^2 is:

$$(811{\cdot}90 - 47{\cdot}99)/10 = 76{\cdot}39$$

The fitted regression line is defined by the equation:

$$\text{output (m.}^3\text{)} = 58{\cdot}2 + 0{\cdot}252 \text{ (shunt flowmeter reading)}$$

The errors in this equation are of two sorts—errors of position, and errors of slope. The variance of the former is the variance of the mean observed output, i.e. $(\sigma_o^2 + 10\sigma_m^2)/200 = 4{\cdot}06$. The first part of this expression—the term $\sigma_o^2/200$—is the variance arising from random errors of measurement only, and would occur even if there were no heterogeneity. The second term ($10\sigma_m^2/200 = \sigma_m^2/20$) represents the additional uncertainty introduced by the heterogeneity; since only twenty separate runs were made, the additional variance of run means is divided only by twenty to give the additional variance of the general mean. It is clear that the heterogeneity is the chief cause of uncertainty about the position of the regression line; $\sigma_m^2/20$ is equal to 3·82, which is 94% of the total variance of 4·06 found above. The variance of the estimated slope also depends on both σ_o^2 and σ_m^2; the term representing random errors of measurement only (i.e. that which would occur even if there were no heterogeneity) is given by σ_o^2 divided by the total sum of squares of the shunt flowmeter readings about their general mean, while the additional variance introduced by the heterogeneity may be shown to be equal to 100 (i.e. 10^2) times σ_m^2, multiplied by the sum of squares of the run mean shunt flowmeter readings about the general mean, and divided by the *square* of the total sum of squares of the shunt flowmeter readings about the general mean. The separate contributions are $0{\cdot}41 \times 10^{-4}$ and $5{\cdot}88 \times 10^{-4}$ respectively, so that once more the uncertainty introduced by the heterogeneity accounts for over 90% of the total variance. Since errors of position and of slope are independent, the variance of the point on the regression line corresponding to a shunt flowmeter reading X is:

$$4{\cdot}06 + 6{\cdot}29 \times 10^{-4}(X - 1143)^2$$

(1143 was the mean shunt flowmeter reading for the series).

We may draw up Table II showing more clearly the errors of fitting.

TABLE II

Errors of fitting

Shunt Flowmeter Reading	Estimated Output ($m.^3$)	Standard Error of Estimate ($m.^3$)	Coefficient of Variation (%)
960	300	5·01	1·7
1,143	346·2	2·01	0·6
1,356	400	5·71	1·4

Over the range of outputs 300–400 $m.^3$/cycle (the range of practical interest) the errors of fitting appear satisfactorily small. Unfortunately, however, they are not the only errors which arise in practice. As was shown by the analysis, the position of the true regression line changes from run to run, and the use of the fitted line to estimate the mean output per cycle for any one run will therefore lead to biased results. Since each run was made on a different day, the run-to-run variation may represent day-to-day variation or shorter period effects. It is in practice advisable to consider it as day-to-day variation since it is commonly found with temporal effects that the variation within days is markedly less than that between days, and there is therefore a serious risk of underestimating the errors possible if we assume that the run-to-run variations observed will average out to any extent in the course of a day. It follows that, if the fitted line is used to estimate the average output per cycle for a single day, the unknown bias must be represented by the variance σ_m^2 found above. If the average output for n days is determined, the unknown bias in the results must be represented by a variance of σ_m^2/n—the day-to-day variation may be assumed sensibly random. This variance must be added to the variance of fitting found earlier. Thus we obtain Table III showing the true errors of prediction.

TABLE III

Errors of prediction

Average Shunt Flowmeter Reading	Estimated Average Output ($m.^3$)	Days Run	Standard Error of Estimate ($m.^3$)	Coefficient of Variation (%)
960	300	1	10·07	3·4
		5	6·35	2·1
		10	5·72	1·9
1,143	346·2	1	8·96	2·6
		5	4·39	1·3
		10	3·42	1·0
1,356	400	1	10·43	2·6
		5	6·93	1·7
		10	6·34	1·6

In the worst case (prediction from a single day's results), errors of up to 7% at the lower end of the range must be expected. Usually, however, results for several days will be averaged, and errors of prediction will then generally be less than 4%, i.e. the effects of the variable daily bias will not be serious. The errors of fitting remain the same no matter how many days' running is considered; this follows from the fact that we are using the same regression line throughout.

In comparative trials using the same calibration curve each estimated output will be subject to similar errors of fitting. These, therefore, can be disregarded in the comparison unless the mean shunt flowmeter readings are different in the different trials. If, for example, two trials each of five days' duration were carried out and the mean shunt flowmeter readings were X_1 and X_2 respectively, the standard error of the difference in estimated mean outputs would be:

$$2 \times (76{\cdot}39/5) + 6{\cdot}29 \times 10^{-4}(X_1 - X_2)^2$$

and not

$$2 \times (76{\cdot}39/5) + 2 \times 4{\cdot}06 + 6{\cdot}29 \times 10^{-4}[(X_1 - 1143)^2 + (X_2 - 1143)^2]$$

The precision of comparison may therefore be rather greater than is suggested by the figures of Table III.

Conclusion

The work reported above formed only a small part of a much more extensive investigation and the results derived are of no great value except in relation to the larger investigation. The analysis was, however, of some interest, since it raised, apart from the perennial problem of which regression line to use, certain points in the treatment of heterogeneous data not previously encountered. It is hoped, therefore, that the account given here will be of assistance to others faced with similar problems, and that it will go some way towards showing that heterogeneity of the data does not, of itself, preclude exact statistical treatment. Heterogeneity is far too common in industrial work to be ignored; the subject would seem to be worthy of a much fuller theoretical treatment than it has so far received.

THE MEASUREMENT OF BRANCH SALES PERFORMANCE

BERNARD T. RAMM

This article shows how figures of turnover have been used to measure the progress of a retail shop in a multiple firm relative to that of the firm as a whole, and how the departments in the shop responsible for any trends are discovered. The methods are illustrated by figures for a firm of retail grocers, but they have a wide application.

Measurement of Total Sales

The simplest and most readily available measure of a shop's sales is the value of the turnover, and at a time of stable economic conditions with steady prices and adequate supplies this figure may perhaps be a reliable indicator of the situation at the branch. These conditions certainly do not exist at the present time, and in recent years large changes in prices and supplies, made worse in the food trade by ration changes and sudden changes in the prices of subsidised foods, have made the value of the turnover by itself useless as a measure of the progress of the sales. Nevertheless value measured in terms of money provides the only convenient means of combining the sales of a large number of dissimilar articles and must form the basis of any overall measure of a shop's trade. It also has the great advantage of being immediately available without extra calculation. However, the measure of the progress of a shop's trade is not necessarily its total value, or even its volume, but the number of customers served and the extent to which their requirements within the range of goods sold are supplied. To investigate progress so regarded, the turnover must be completely analysed; changes in its level must be related to changes in prices and in the sales of individual items; and an attempt must be made to explain these changes. The sales figures must be examined against the background of such national statistics as are available, the results of consumer surveys, and, in the case of the food trade, ration levels and registration figures. This is a somewhat elaborate investigation, and although it is important that it should be carried out at regular intervals on the total trade of the firm, it is obviously out of the question to make it for each of several hundred branches. Fortunately this is not necessary; once the progress of the firm as a whole has been determined in this way, all that is required is a ready means of comparing the sales at each branch with those of the whole firm. This comparison can best be made in three stages: first a simple overall comparison, which must quickly spotlight any branches whose progress is markedly different from the average; second a comparison of the

sales divided into departments, to show broadly where this divergence lies; and third an examination of the sales of individual lines. This last cannot be made a routine examination owing to the amount of work involved and should only be made when no ready explanation can be found for divergences shown up by the first two comparisions.

The most obvious overall comparison is one between the percentage change in a branch's takings over a period and that for the whole firm over the same period. This comparison may, however, be very misleading. The different departments of the business are not likely to be making the same progress; for instance, in the food trade an increase in the size of the rations is normally accompanied by a decrease in the demand for other foodstuffs such as groceries and canned goods. Moreover different branches will not have their trade divided among the various departments in the same way, for even if all branches sold the same range of goods and were treated in exactly the same way local competition and tastes would soon cause differences in the make-up of their trade. Such differences will mean that even if each section of a branch's trade is following the average trend perfectly its total takings may appear to be moving quite differently from the average. Suppose, for example, that throughout a multiple grocery business the turnover of rationed foods has increased by 50 per cent. during a certain year and of unrationed foods by only 10 per cent. Then a shop with a trade at the start of the year of £300 a week in rationed foods and £500 in unrationed foods would be expected to show a 25 per cent. increase in its turnover, whereas one with a trade of £300 a week in rationed foods and only £300 in unrationed foods should show an increase of 30 per cent. Suppose now that at the latter branch an increasing number of registered customers were going elsewhere for their groceries and the trade in unrationed foods only increased by 5 per cent. The increase in takings would be $27\frac{1}{2}$ per cent., a figure still above average although the shop was losing custom. These variations in the make-up of a shop's trade are not greater than those that occur in practice, and they must be taken into account before any comparison can be made; in fact a still finer subdivision than that between rationed and unrationed commodities is necessary, depending on the business.

Another desideratum of the comparison that is not fulfilled by this simple method is that it should provide a continuous record of each branch's progress and not merely a comparison between the trade at two particular dates.

A solution of the problem is to compare each month's takings with an expected value derived from the branch's actual trade in a base year, assuming an average progress in each department. The calculation of this expected value is illustrated by Table I. Column (*a*) gives the branch's average four-weekly turnover during the base year and column (*b*) the whole firm's turnover in the month under consideration expressed as a percentage of its average in the base year. These percentages are then applied to column (*a*) to give the expected values in

column (*c*), the total of which is then compared with the actual takings for the branch. The result of that month's trading is thus summarised in one single figure: the percentage difference between the expected and actual values of the turnover.

TABLE I

(a) *Average four-weekly turnover at Branch X for year ending 31st March, 1950 (base year).* (b) *Turnover for the whole firm in four weeks ending 7th October, 1950, as a percentage of the average for the base year.* (c) *Expected turnover at Branch X in four weeks ending 7th October, 1950.*

Department	(*a*)	(*b*)	(*c*)
	£	%	£
Rationed trade	1,000	150	1,500
Groceries	500	110	550
Canned goods	700	130	910
Preserves	100	100	100
Cooked meats	500	150	750
Poultry, game, and rabbits ..	400	80	320
Caterer's groceries	200	110	220
Total turnover at Branch X in four weeks ending 7th October, 1950 — Expected			4,350
Actual			4,000
Difference			− 350 = − 8%

An illustration of the way these figures run is given by the first line of Table II, which is based on the performance of an actual branch. It is immediately obvious from these figures that this branch has not made average progress, losing considerable ground between April and August, and since the middle of July the takings have been 11 per cent. below the level to be expected from the trade in the base year.

A difference shown by this measure need not be indicative of something peculiar to that branch alone, but may be due to different seasonal variations in certain areas, e.g. seaside resorts as an extreme case, or a change in local economic conditions. This factor can be dealt with by averaging groups of shops to obtain regional values and then comparing the figures for individual branches against these. Besides revealing branches or areas where the sales are falling below average the figures also provide an immediate estimate of what the effect on the takings has been of any alteration to a branch, such as expansion of the premises, conversion to self-service, or a change in local competition.

Sales in Different Departments

The next stage after this examination of total sales is to determine which section of the trade is responsible for any abnormal trend. If a departmental division of each branch's takings is made available by the use of analysing cash registers, the problem is simple. All that is required is to express the trade in each department during each four-weekly period as a percentage of that department's sales over the whole

firm. A steady change in these percentages for any department will then show straight away that its progress is differing from the average for the firm. Such a division of the takings, however, is not usually available, and the departmental figures may have to be arrived at indirectly. If branch stocks are strictly controlled these departmental percentages may, with little loss in accuracy, be derived from the value of goods supplied to the branches instead of the goods sold. If the goods are charged to the branches by means of a punched-card accounting system it is a simple matter to obtain a tabulation by department of the goods supplied to each branch and to use this to calculate the percentage of the whole firm figures. Another alternative would be to obtain sales figures for a number of representative lines and to use these as an index of the progress of the whole of the department to which they belong. The difficulty here lies in choosing the lines to be used in the index, for it is not easy to find items which are representative for every branch and yet few enough in number to keep the work reasonably low.

Now we return to Table II. The next step after the overall comparison is to examine the departmental figures, the first and most important of which are those for rations. Apart from slightly low figures in April (when eggs were plentiful and there was not the same demand for them at this branch (a country shop) as in the town) and during the holiday season, rationed sales have remained steady, and this can be confirmed by the registration figures. The next line shows the relative level of the grocery sales, and here there has been a large fall—from 0·58 per cent. in April to about 0·46 per cent. during the last three months, a drop of a quarter in the relative level of the sales. Canned goods sales are somewhat erratic owing to allocations of scarce items, but do not show a downward tendency. The preserves figures are also erratic, showing the effect of special 'pushes' and offers of cheap jam. The level of cooked meats sales rose to a peak in June and July and subsequently fell slightly, but the variation here has been slight. The poultry, game and rabbits sales show a 25 per cent. drop from their level in April and May, very similar to the drop in the grocery department. Caterers' sales vary considerably from month to month, but there is no obvious trend.

It is therefore clear from these figures that the fall in the overall level of the trade in the branch lies in the grocery and poultry, game and rabbits departments. The next step is an enquiry at the branch in an endeavour to find an explanation—a clue could possibly be given by the very high level of the grocery sales relative to the rationed sales in April (0·58 per cent. against 0·32 per cent.). If no satisfactory explanation can be found, the sales of a number of individual lines would have to be examined, with a view both to showing more exactly where the drop in sales has occurred, and to finding out in what items grocery sales were above the average and why these later fell to a more normal level.

TABLE II

Branch sales record, Branch No. 514

1950 Four weeks ending:	22nd April	20th May	17th June	15th July	12th Aug.	9th Sept.	7th Oct.	4th Nov.	2nd Dec.
Percentage difference between actual and expected takings	− 2	− 5	− 6	− 6	− 11	− 12	− 11	− 11	− 9
Department	Takings at branch as percentage of takings in department for the whole firm								
Rationed food ..	0·32	0·34	0·35	0·36	0·33	0·33	0·35	0·34	0·35
Groceries	0·58	0·54	0·53	0·50	0·39	0·50	0·45	0·47	0·48
Canned goods ..	0·32	0·46	0·42	0·45	0·41	0·46	0·48	0·44	0·33
Preserves	0·51	0·53	0·60	0·49	0·73	0·61	0·33	0·58	0·38
Cooked meats ..	0·38	0·40	0·41	0·41	0·39	0·40	0·39	0·39	0·36
Poultry, game and rabbits	0·41	0·38	0·28	0·34	0·31	0·27	0·30	0·30	0·33
Caterer's groceries and canned goods	1·11	0·69	0·82	0·85	0·45	0·63	0·68	0·82	0·93

The departmental percentage figures have many uses other than that of determining the trend of sales. If the total sales of a particular line are felt to be insufficient, one of the first steps towards increasing them is to find out where the sales are low, and these departmental percentages form a valuable guide as to the quantities each branch should sell. In this type of investigation it is important to take into consideration the type of customer at the branch and to make comparisons between similar shops. For example, we find that in working-class areas the most important factor in determining the quantity of tea sold is the number of sugar registrations, whereas in residential areas it is determined by the amount of grocery trade done by the shop. The figures also lend themselves very readily to the conversion of a sales forecast for the whole firm into 'target' figures for individual shops, and they can also form a basis for the allocation of scarce goods.

Besides measuring each branch's share of the firm's trade and the way in which it is changing with time, the figures provide a very convenient means of comparing the levels of the sales in different departments at the same shop. In the completely average shop all the departmental percentages will be equal and the ratios between them will give a measure of the constitution of each branch's turnover. The variations from equality will in most cases be readily explicable in terms of factors such as the type of customer, space limitations, or local competition; but a valuable clue can sometimes be given as to departments in need of development.

ONE LINE OR TWO?

WILLIAM N. JESSOP

Operational Research Department, Courtaulds Ltd.

Many an experimenter has tried to use the method of least squares to estimate the true straight-line relationship between x and y when he has a few observed values subject to error, and has been exasperated to learn that the method gives two lines. Mr. Jessop surveys recent theoretical work that has been done on the problem and shows how a single line may be determined. He also discusses in what practical situations it is appropriate to use the one- and two-line relationships.

Introduction

Those familiar with statistical methods often do not appreciate that the representation of a relationship by two regression lines is mystifying to many technologists who think in terms of single formulae which give one of the related quantities knowing the values of the other. The statistician explains that one line is to be used to estimate y knowing x, and the other to estimate x knowing y. It would not be surprising if the technologist attempted a representation in accordance with his own experience by drawing a line midway between the regression lines, with the intuitive idea that a single relationship exists. When the statistician sets to work to show the essential unsoundness of this procedure he will be led to a more critical appraisal of his own, and will at least have to consider whether there are circumstances in which single-line representations have a meaning. This article is concerned with clarifying this situation by reference to the results of theoretical papers: results which have not yet been incorporated in the textbooks.

Classical Case

The word regression itself was first used in connection with the following biological example. The data consist of pairs of measurements of 'height of father' and 'height of son.' Then, calling the former x and the latter y, two distinct lines are obtained. Each regression line serves the purpose of enabling the mean value of one measurement to be found if the other is given. The question of a unique 'true' relationship does not arise; we have no body of theory, as we might have were the study physical, to assert that a functional relationship $y = f(x)$ should exist and that the scatter of points on the graph is the result of observational error; the purpose of the two lines is to describe the characteristics of a bivariate normal population in so far as they can be judged from a sample. No single line could tell us as much as the two lines do.

Functional Relationships

In the preceding, error is very unimportant compared with biological variation; it is the latter that was minimised to get the best estimate of the mean y for a given x and the mean x for a given y. However, when our data are physical measurements biological variation does not enter the picture, and it is in such situations, among others, that the question of a single-line representation does arise. We may, for example, know from theory that the law governing the relation between y and x is of the form $y = ax + b$, and our interest lies in determining the values of a and b. The usual approach is to make the assumption that x is free from error and determine the line to minimise the sum of the squared deviations of y, this representing error in y. This method would be sound applied to the case of a density determination where volumes (y) were measured with error and masses (x) with great precision. Often the assumption that one of the variables is free from error is highly questionable; and if in this example masses are not measured with high accuracy, what is required is a method which will estimate a constant (density) in a functional relationship, account being taken of the errors in the two measurements.

One of the earliest attempts at a method of arriving at a linear functional relationship, allowance being made for errors in both variates, is that due to Kummell (1879). In the light of modern theory it is a most successful attempt, arriving at results similar to those obtained recently by Mr. Lindley.[3] Unfortunately Kummell's work was published in an obscure journal and for many years was overlooked.

Better known is the method of Karl Pearson (1901), which is to minimise the sum of the squared deviations measured in a direction at right angles to the line rather than parallel to either axis as in the usual case. C. F. Roos[4] points out, in a paper which is particularly valuable as a summary of the work on this subject up to the year 1937, that this method is open to the serious objection that changing the scale of either x or y will alter the line so obtained. Thus if x were mass, a different line would be obtained by changing from tons to pounds or kilograms; and similarly for y. The method of fitting lines 'by eye,' which roughly approximates to this method, is open to the same objection.

A paper which does much to clarify the problem is that of Mr. Lindley.[3] He shows that a satisfactory estimate, in the light of modern estimation theory, of a linear functional relationship $y = ax + b$, may be obtained if the relative magnitudes of the errors in x and y are known. The formulae to be used are as follows:

$$a = m + \sqrt{(m^2 + k)}$$
$$b = \bar{y} - a\bar{x}$$

where $\bar{x}$ = average of the x observations,
$\bar{y}$ = average of the y observations,

$$k = \frac{\text{error variance of a single } y \text{ observation}}{\text{error variance of a single } x \text{ observation}} = \frac{v_y}{v_x}$$

and $$m = \frac{s_{yy} - ks_{xx}}{2s_{xy}}$$

where s_{xx}, s_{yy}, and s_{xy} are the sums of squares and products generally used in the calculation of correlation and regression coefficients.

Consider the following artificial example. In an experiment to determine the elastic constant of a material the following values of stress (x) and strain (y) are obtained.

Stress (x) ..	1·1	2·2	2·9	4·1	5·0
Strain (y) ..	0·5	2·5	2·5	4·5	4·8

This gives $s_{xx} = 9{\cdot}45$, $s_{xy} = 10{\cdot}46$, and $s_{yy} = 12{\cdot}23$. If the observations of x were made with very high accuracy as compared with those of y the experimenter would take the regression of y on x as the best means of determining the elastic constant, i.e. $1/a$. On the other hand, if the strain observations (y) had been made with higher accuracy than the corresponding stress observations (x), he would have chosen the other regression line for his estimate. The two cases correspond to $a = 1{\cdot}11$ and $1/0{\cdot}86 = 1{\cdot}17$. These determinations are, in fact, the same as those which would be obtained by Mr. Lindley's formulae, since these reduce to the ordinary regression slopes when one or other of the quantities x and y is accurately known and the other subject to error.

But what of the case where neither set of observations is perfect? Provided we know their relative accuracy, i.e. the value of k, the formulae above provide an answer. For example, suppose we know that in the units chosen the error variances of the stress and strain measurements are equal, or $k = 1$.

Then $$m = \tfrac{1}{2}(12{\cdot}23 - 9{\cdot}45)/10{\cdot}46 = 0{\cdot}13$$

so that $$a = 0{\cdot}13 + \sqrt{(0{\cdot}13^2 + 1)} = 1{\cdot}14.$$

This lies midway between the two regression lines.

If the value of the ratio k is not known but one of the error variances is, then a may still be estimated. Mr. Lindley gives the following formulae:

$$a = \frac{s_{yy} - nv_y}{s_{xy}}$$

$$v_x = \frac{as_{xx} - s_{xy}}{na}$$

where n is the number of observations.

Thus if we know v_y we can estimate a and the error variance of the observations of x. Suppose, in the above example, strain is known 'with an accuracy' of $\pm$ 0·28 unit; and further, suppose that this statement has a precision which it often lacks, so that the 0·28 is twice the standard deviation. Then, $v_y = 0{\cdot}02$ and

$$a = \frac{12{\cdot}23 - 0{\cdot}10}{10{\cdot}46} = 1{\cdot}16$$

$$v_x = \frac{1{\cdot}16 \times 9{\cdot}45 - 10{\cdot}46}{5 \times 1{\cdot}16} = 0{\cdot}09$$

i.e. the 'accuracy' of a stress observation is $\pm$ 0·6 unit.

The value of k is 0·02/0·09 = 0·22, and it may be verified that the same value of a is obtained by employing the previous formula and using this value of k.

Effect of Manner in which Observations are Obtained

Dr. C. Eisenhart[2] has drawn attention to the importance of a critical study of the data, since a mechanical application of regression analysis can result in misleading conclusions. Suppose that of the two variables x and y, x can be held at preselected values whereas y cannot; then in the resulting data it is the observed y which is in error, so even if we are interested in using observed values of y to estimate x we should still fit $y = ax + b$ and use the inverse, i.e. $x = (y - b)/a$, to estimate x. Further, although it is legitimate to select values of the independent variable (x in the above), it is not permissible to do so for the dependent variable; to do so could seriously distort the relationship. These remarks apply to both of the cases we have distinguished: the biological bivariate population and the observations having a functional relationship.

Dr. J. Berkson[1] continues this line of thought a stage further. He discusses the notion of an 'observation,' showing that two distinct types arise in practice. In the first case we have what he calls an 'uncontrolled observation' which could be a physical measurement of some sort, e.g. a weight, length, or period of time; the observations are such that the mean of a series of them measured on a single variable has the true value as its mean. It corresponds to what we usually consider to be an observation. The second case is called a 'controlled observation.' This is made when, instead of determining some unknown, we bring the quantity under consideration up to a preset value. Dr. Berkson cites the case of quantitative chemical analysis in which it is required to add specified weights of a test material to a solution, and bio-assay experiments in which organisms are exposed to increasing specified concentrations of the substance to be assayed. In this case the mean of the true values has as its mean the preset value. For example, suppose the preset value is 8·00. Then if we carry out five

controlled observations (by adjusting conditions such that each is 8·00 nominally) the mean of the true values of these five may be 7·95, and for another set of five the mean may be 8·06; but if we continued this process indefinitely the mean of these values would be 8·00.

After making this distinction Dr. Berkson derives the following results:

'(i) The line estimated by least squares, minimising the sum of the squared residuals of the dependent uncontrolled variate, is the same on the average whether x or y is the controlled variate; that is, there is only one true regression.

'(ii) The estimated line is not biased by the existence of an error of observation in the independent controlled variate, despite no account being taken of it in the least squares fit.'

It should perhaps be remarked that (i) does not imply that the two observed regressions would be the same. The difference, however, would not be real; it would reflect only sampling variation. In regard to (ii) it can be established that if the independent variate is uncontrolled the existence of errors in its values tends to yield results which understate the true relationship. This happens when we simply collect records of simultaneous pairs of values. We have then samples from a population of observations which are uncontrolled. Thus, to use Dr. Berkson's own example, if we are investigating an Ohm's law relationship $V = RI$, we may collect simultaneous readings of voltage V and current I in order to determine R; then if V is subject to an unbiased error of measurement the regression of I on V will on the average overestimate R, i.e. underestimate the slope $1/R$. On the other hand we may measure the current I for different specified values of the voltage V, or alternatively the voltage V for different specified values of the current I. Each is a controlled experiment in the sense that in one case values of V, and in the other values of I, are controlled observations. Then by (i) and (ii) the two regressions obtained in these cases are estimates of the same regression, and further, the values of R derived will not tend to overestimate R as in the result obtained from data which 'just happen.'

A question which arises is: if practical convenience does not determine which is to be the independent variate, what other criterion should we use ? From a consideration of the standard errors of the two estimates of slope the rule may be given to choose as dependent variate the one which is likely to give the larger error when considered as a controlled variate. When practical convenience does enter the picture, this will often mean that the best choice is made anyway, since a variate which is difficult to bring to a specified value will frequently be subject to the larger error.

Dr. Berkson refers to another case which he describes as the 'modified controlled experiment.' Here one of the variates is brought into the neighbourhood of a series of specified values. This may be illustrated by the data of the stress-strain experiment where the target stress figures

are $x = 1, 2, 3, 4$ and 5. The resulting regression tends to give too low a value for the slope of strain on stress. However, if semi-controlled observations are taken over a wide range this bias can be reduced to a negligible amount. This case is worth mentioning, since many experiments are of this type.

It should be understood that the one-line regression case applies whether or not the relationship is functional, the condition residing in the way the observations are obtained and not in their nature. If a sample is taken, we have two lines; if one of the variates is made to take a desired set of values, we have one.

Concerning Two Lines

A great number of statistical enquiries are concerned with the estimation of one variate given the other, and here the two regressions retain their usefulness whether the variates are functionally related or not. If x and y are functionally related the two regression lines calculated from the observed values are such that for a given observed x we can estimate the true value of y and *vice versa*. If x and y are not functionally related but are such that the regression of one on the other is linear, either regression line enables the true mean value of one to be estimated from a given observed value of the other.[3]

Assumptions

At all stages in the preceding discussion it is assumed that errors are random and have a mean of zero when due regard is taken of sign, i.e. the observations are made with unbiased error. The two types of observations distinguished by Dr. Berkson differ in that an uncontrolled observation has an error which is independent of the *true* value, while a controlled observation has an error which is independent of the *observed* value. Mr. Lindley's equations for the estimation of a linear functional relationship further assume that the errors are normally distributed about the true values.

Practical Upshot

Many users of statistics, in fields removed from laboratory research, will find the classical two-line analysis most appropriate, since they merely require the best estimate of one variate given the other. The data will be of a sort requiring a statistical rather than a functional description, so that unless the data have been obtained by controlling one variate in the Berkson sense, the question of one line does not arise.

Other workers will be interested in scientific laws and the determination of the constants in them, due allowance being made for observational error. It is to this group that the usual two-line representation is open to objections. The types of experimental situation are so varied that a few general remarks can hardly be satisfactory. However, we can suppose two contrasted situations: (1) where the experimenter has

no control of the values taken by the variates being measured, or circumstances oblige him to accept what happens, and (2) where the experimenter can control at least one of the factors. The first type of situation may arise outside the laboratory, for example in studying the operation of some unit of industrial plant. This is a case for using Mr. Lindley's formulae provided we have knowledge of the errors to which the observations are subject. In this connection one hears of observations being accurate to within so much per cent., but it is often not clear as to how such assessments are made or what precisely is meant by them. It may be necessary to carry out a separate experiment to get a reliable idea of observational error. For example, it may happen that the experimental circumstances do not permit the use of high-accuracy instruments. In this case it is suggested that the instruments to be used in the experiment should, if possible, be compared with more precise ones in a statistically designed trial and the errors of the former assessed.

In the second situation we may still take observations at random and apply Mr. Lindley's results when accurate knowledge of errors exists. A better method, however, may be to arrange for one of the variates to be 'controlled.' The regression of the uncontrolled on the controlled observation then estimates the functional relationship, and we need not trouble about errors in making this calculation. This procedure is one commonly practised, so that Dr. Berkson's paper is reassuring.

REFERENCES

[1] Berkson, J. (1950). 'Are There Two Regressions?' *J. Amer. Statist. Assoc.*, **45**, 250.

[2] Eisenhart, C. (1939). 'The Interpretation of Certain Regression Methods and their Use in Biological and Industrial Research,' *Ann. Math. Statist.*, **10**, 162.

[3] Lindley, D. V. (1947). 'Regression Lines and the Linear Functional Relationship,' *J. R. Statist. Soc. Supp.*, **9**, 218.

[4] Roos, C. F. (1937). 'A General Invariant Criterion of Fit for Lines and Planes when all Variates are subject to Error,' *Metron*, **13**, 3.

References not mentioned in text

Bartlett, M. S. (1948). 'The Fitting of Straight Lines if Both Variables are subject to Error,' *Biometrics*, **5**, 207.

Deming, W. E. (1943). *Statistical Adjustment of Data.* Wiley, New York.

Kendall, M. G. (1951). 'Regression, Structure and Functional Relationship,' *Biometrika*, **38**, 11.

Wald, A. (1940). 'The Fitting of Straight Lines if Both Variables are subject to Error,' *Ann. Math. Statist.*, **2**, 284.

NOTES AND COMMENTS

THE Teaching of Statistics in Schools is the title of a report of the Council of the Royal Statistical Society, recently published in Part I of the 1952 volume of the society's *Journal, Series A*. The report argues that all children, including those who attend modern secondary schools, should learn something of statistical techniques and the 'statistical approach' in order to enable them to play their part as citizens in a democracy. In addition, it is argued, statistics should find its place in the more specialist studies of secondary and technical schools because of the increasing importance of its applications to business and administration, engineering and industry, and science and economics. School is the place to learn statistics, not only because most children finish their formal education there, but also because statistical methods and ideas are often found to be difficult—not that they are intrinsically so but that people find it difficult to pick up a new set of ideas after having become habituated to an old set. A comparatively leisurely introduction to statistics at school makes it more likely that the more intensive courses at the university will be assimilated.

The report mentions detailed considerations involved in carrying out the recommendations but leaves it to educationists to deal with them. Difficulties are admitted, but the Council of the RSS believes that 'if the case for statistics is examined on its intrinsic merits and is not prejudiced by the newness of the subject and lack of understanding, the difficulties will be overcome and statistics will find a proper place in the general and special curricula.'

★ ★ ★

Two new local groups of the Industrial Applications Section of the RSS have been formed in recent months. The Merseyside Group was inaugurated at a meeting of 120 people held on 31st January, 1952, in Liverpool. The chairman was Mr. O. P. Perrin of Messrs. Dunlop Ltd. and the meeting was addressed by the Lord Mayor of Liverpool, Mr. Philip Lyle, Dr. B. P. Dudding, and Mr. O. P. Perrin who was also organising secretary. The group will function informally for a short period, and several meetings have already been held. The local secretary is Mr. R. L. Plackett of the University.

The Leicester Group was inaugurated on 1st April, 1952, at a meeting arranged by Mr. M. J. Moroney in the Leicester College of Technology. About fifty people attended. Mr. K. Boyd of British Timken Ltd. outlined the functions of a local group of the Industrial Applications Section, and Miss J. Keen, honorary secretary of the Section, dealt briefly with the history of the Section and the Society.

A local committee was formed with Mr. Boyd as chairman and Mr. D. Bramley of B. B. Chemicals Ltd. as honorary secretary.

★ ★ ★

There are relatively inexpensive card index systems that use cards on which data in a coded form are punched by a hand instrument, and in which classes of cards can be readily selected from a stack with a simple device incorporating a long selecting needle. Although apparently not designed for such purposes, these systems have been found useful in statistical work. Their main disadvantage is the time taken to count the cards after sorting. Dr. B. M. Wright of the Pneumoconiosis Research Unit, Llandough Hospital, near Penarth, Cardiff, has devised and had made an automatic device for counting which is designed in detail to deal with 'Findex' cards, although it could be made to deal with any cards of somewhat similar shape and size—their thickness, particularly, needs to be round about one-hundredth of an inch. The mechanism does not depend on the cards being more uniform in thickness and weight than such cards usually are, and it has been found to be very reliable in operation. Dr. Wright writes:

'The machine separates and counts the cards at a rate of 500 per minute, being limited by the speed of the electromagnetic counter. Most of the time taken by the person using it is spent in straightening up the cards and placing them on the machine. A moderately skilled person can count 60 cards on the machine in about 20 seconds, of which only 10 seconds is actual counting time. A highly skilled person can do the same task by hand in 60 to 90 seconds, though the average person would take at least twice as long as this. The time saved is therefore considerable, but the real virtue of the machine is that it enables counting to be done by an unskilled person who is not paying much attention to what he is doing. The machine has been in constant use for over two years, during which time it has counted over 1,000,000 cards. It is of relatively simple design, based on kinematic construction, so that what little wear takes place is automatically compensated. Most of the breakdowns that have occurred have been in the enumerator, as no suitable electromagnetic counter appears to be available commercially.'

Dr. Wright has attempted to interest firms in the commercial development of the device, but so far without success; apparently the anticipated demand is not sufficient. He would be glad, however, to show the machine to anyone who is interested in the use of card systems for statistical work.

INDUSTRIAL APPLICATIONS SECTION OF THE ROYAL STATISTICAL SOCIETY

Reports of Meetings

BIRMINGHAM AND DISTRICT GROUP

26th September, 1951. Chairman's Address—The Precision of Statistical Tests: D. G. Beech

Dr. Beech discussed some of the principles underlying commonly used significance tests—particularly Student's *t*-test and quality control charts. He stressed that although errors of the first kind were always considered, the importance and magnitude of errors of the second kind were frequently ignored. He described some of his own investigations into the examination of trends in quality control charts—an approach often advocated but seldom treated in detail, suggesting the use of an additional warning limit so chosen that the chance of two consecutive points falling beyond it is, say, one in 40. If two consecutive points, two out of three, or three out of four, lay beyond this limit they were pooled and tested provided the remaining point were near the limit. In spite of its subjective flavour such a test had proved to have greater power for detecting changes in the system than more usual tests. Longer trials were now necessary to establish its value.

24th October, 1951. Statistics in Trouble-shooting: E. A. G. Knowles

Dr. Knowles gave two examples to show how elementary statistical methods can be used to eliminate or reduce production difficulties. The first arose during the manufacture of electronic valves. The fragile coating of the heater could be damaged during assembly and processing, making the valve defective. The proportion so lost varied from time to time. Judgment of the fragility of heater coatings was entirely subjective and open to disagreement. A simple impact test was designed in which a glass rod fell from a set angle on to the heater. This gave a measure of 'toughness' which seemed to be normally distributed among good heaters; but among heaters of rejectable quality the distribution was skew, with lower mean and less dispersion. A control chart showed the differences between good and rejectable quality to be so striking that a 'go—no go' method of testing should be sufficient. This was arranged so that in the long run about 10 per cent. of the good-quality heaters would fail the test. Control charts showed that the nature of the differences between good and bad heaters varied and gave indications of likely causes of trouble.

For her second example Dr. Knowles described how, apparently owing to processing errors, the bulbs of cathode-ray tubes had 'imploded' during processing. How should the specification be improved to reduce danger of 'implosion'? Increasing the severity of the pressure test gave more test breakages, so that frequency diagrams could be drawn of various measurements made on the bulbs and the corresponding proportions of breakages, and χ^2 tests on these data showed that one type of bulb could be made

stronger by increasing the screen thickness and another by raising the specified minimum weight of the bulb. Further experiments confirmed the soundness of these conclusions.

2nd January, 1952. The Statistical Approach to Time Study: D. J. Desmond

After briefly surveying time study, Mr. Desmond described a new method for analysing time study data: the 'Reciprate Method' based on regression analysis, which gives an objective estimate of the normal time of a job in terms of the recorded times and subjective ratings of the observer. The precision of this estimate can be calculated and comparisons can be made between observers studying the same job. The various defects in a study can be calculated in terms of three parameters which establish the standard of quality of the observer's studies. Plotting these statistics on control charts enables the observer to determine readily whether he is maintaining the quality of his work and to see if he is making any improvement. A simple graphical method was described which enables estimation of all the characteristics of a study in less time than was usually taken to determine the normal time. Mr. Desmond then described how, by the analysis of variance technique, the results for any number of studies could be combined and how this led to the establishment of a standard of quality for a group of observers and to the examination of the significance of differences between observers. These differences were illustrated by the results of an experiment carried out on the floor of an assembly shop.

LONDON GROUP

2nd November, 1951. Making Practical Decisions on Statistical Data: G. A. Barnard

The paper consisted of a description of the main features of the Theory of Decision Functions as developed by Professor Abraham Wald, and a statement of Professor Barnard's own views. It is assumed in Decision Theory that measures of 'damage' are available for each possible action in each possible situation. Further, there are observations whose probabilities, being known in each of the possible situations, give evidence as to which of the situations is the true one. The cost of making these observations must also be known. The total cost involved in adopting a given decision rule (i.e. the sum of the damage and the cost of observations) is thus associated with a probability so that an expected loss can be calculated. It would be on the basis of these expected losses that decision rules would be chosen. A choice based on minimum expected loss is seldom possible. The best rule will depend on which situation is the true one and then the 'minimax' principle is favoured, at least in the USA. The minimax principle chooses the decision function to give the least maximum loss. Professor Barnard regarded it as unduly pessimistic to base one's actions on the worst possible consequences.

7th December, 1951. The Planning of Experiments in the Chemical Industry: G. E. P. Box. Joint meeting with the Society of Chemical Industry

Mr. Box treated in a more general way the subject of an earlier paper written by him jointly with Mr. K. B. Wilson (*J. Roy. Statist. Soc. B*, **13**,

1951, p. 1). Most experiments in the chemical industry had as their object the choice of levels of various factors to maximise some property such as yield, purity, or cost. If the results of an experiment are represented geometrically, using the factors and the property as co-ordinates, the resulting points lie on a surface and the object of further experiments will be to determine the high points on this surface. Mr. Box had shown that it will be best to proceed along the line of greatest slope. He pointed out that the high points might appear either as points, or a ridge, or a rising ridge, or a col. He showed how each type could be identified by reducing the expression representing the fitted surface to a standard form. The maxima were in practice very often of the ridge type, thus giving great scope for the choice of actual process conditions. The talk was illustrated with actual examples.

NORTH EASTERN GROUP

17th October, 1951. Training Personnel in Quality Control: N. D. Squirrell

One of the main things liable to cause a quality control system to break down is the lack of knowledge of statistical method among inspection and production staffs of the lower wage groups. Mr. Squirrell's own firm had endeavoured to overcome this by a series of five short lectures to all members of the staff who were likely to come in contact with quality control charts. These lectures covered the normal distribution, the derivation of means, the setting of control limits, and so on.

21st November, 1951. The Significance of Accident Records: H. Campbell

In the field of industrial health there are many variables. Different industries have different accident rates, and the type and severity of accidents varies: from one industry to another and often between factories in a group. Much valuable information can be gained from a study of the reports of the Inspector of Factories; for example, accident rates in similar industries and factories can be compared. Mr. Campbell defined different types of accidents and illustrated graphically the severity frequency distribution, severity being measured by time lost from work because of an accident. Graphs designed to show actual against expected accident rates, if fitted with control limits, would reveal any significant change, and analysis of this data might indicate steps to be taken to prevent increases in accident rates. Mr. Campbell concluded with a discussion of accident proneness, which produced the expression of some varied opinions from those present.

4th December, 1951. The Measurement of Productivity: D. J. Desmond. Joint meeting with the Association of Cost and Works Accountants

Mr. Desmond gave various definitions of productivity, which could be of either men or machines. He then discussed time and motion study, giving comparisons between different time study engineers and also between firms making similar products. Factors affecting the quality of time study were the engineers' varied conceptions of what constituted normal performance, and the 'flatness' or 'steepness' of rating. Mr. Desmond showed, with the aid of slides, how average performances could be arrived at, and by graphing

results of observations and fitting control limits significant variations could be noted. Management indices and the statistical significance of monthly variations were also explained.

16th January, 1952. The Rational Planning of Experiments: G. E. P. Box. Joint meeting with the Society of Chemical Industry

This address was substantially that given to the London Group on 7th December, 1951.

20th February, 1952. Statistics and Management: D. Gold

A statistical department could be of great value to management in formulating plans and schedules, and in showing whether these are proceeding correctly. Mr. Gold thought that the statistician in an industrial field must have an industrial background. Although simple arithmetic answers most management questions, when the 'how and why' is asked the situation is more difficult and the statistical approach more necessary. Mr. Gold then dealt with the summarising of large numbers of returns, a subject involving problems of form design and of recording. Any estimates given should show upper and lower limits. Data relating to costs and to productivity are an important requirement.

19th March, 1952. The Arithmetic of Statistics: D. Gold

Mr. Gold first explained the concept of a distribution, giving examples of distributions encountered in practice as well as theoretical ones. He went on to explain the measurements used in specifying the characteristics of distributions and the idea of statistical significance, producing charts and graphs and inviting questions on them. From the ready response of the audience it was apparent that there is a demand for further lectures on such fundamental subjects.

SHEFFIELD GROUP

27th September, 1951. Sidelights on the Error Curve: J. F. Hinsley

In the course of an interesting and stimulating address, Mr. Hinsley gave examples of frequency distributions from everyday life, including the distribution of the widths of tree rings, of wear at a drinking fountain due to a falling cup, and of the wear on a short flight of steps. The number of events causing wear was large, approaching infinity, and the shape of a step after wear was shown to agree well with that of the normal curve. Mr. Hinsley dealt with the relationship between the Poisson, binomial, and normal distributions and emphasised the utility of transformation for producing variates that are distributed normally. He described the Herne method for obtaining a rapid experimental solution to a situation involving random events and its application to a stockpiling problem.

1st November, 1951. A Comparison of Different Methods of Inspection: B. H. P. Rivett

In discussing the main types of conventional sampling scheme Mr. Rivett introduced the idea of the Process Operating Characteristic (POC), which

describes the performance of a process, and the Inspection Power Curve (IPC), which characterises the method of inspection and is independent of the POC. The object of inspection is to improve the POC, and the efficiency of an inspection scheme can be assessed by comparing the original POC without inspection and the final POC after the operation of the scheme. It is usually relatively easy to determine the IPC once the scheme has been chosen, but difficult to design a most efficient inspection scheme to produce a pre-assigned POC. Mr. Rivett described rational sentencing schemes as being among the more powerful inspection schemes, and mentioned modifications designed to avoid the sacrifice of lots delivered before or after a known change in process has occurred.

28th November, 1951. Elementary Statistical Methods as Aids to Foundry Production Control: W. G. A. Jenkins

Mr. Jenkins gave a full and interesting description of how a new and successful method of production control had been developed in a foundry. The statistical part of the work involved the analysis of foundry records by frequency diagrams and control charts, and this led to the establishment of new planning criteria.

31st January, 1952. Some Applications of Sequential Sampling to Steel Works Problems: G. F. Komlosy

After reviewing the principles of sequential sampling as first enunciated by Wald, Mr. Komlosy sketched a method for determining the operating characteristic and average number of observations for a test which examines whether the mean of a normal distribution of known variance falls short of a specified value. He described the application of the techniques to the check calibration of a carbometer, showing how the method of checking the calibration was modified as a result of the tests. Mr. Komlosy also discussed the use of sequential analysis in interpreting the significance of 'runs' in control chart work.

28th February, 1952. The Use of Experimental Designs in University Technological Research: N. L. Franklin

Mr. Franklin discussed two main topics: the type of course presented to students, and the application of particular types of design to technological research in chemistry.

Since it was not possible to include an initial course in statistics in the syllabus, his object was to inform the chemical engineer of the statistical techniques available, to indicate how efficient sampling schemes could be developed, and to familiarise the student with simple factorial designs. The graduate student had a wide range of interests, from the purely scientific to the technological. A main interest lay in the sampling of raw materials and in the investigation of errors in a measurement process. To meet the requirements without going into mathematical detail, examples of typical testing schemes and designs were discussed in the course, and help was given to individual research students faced with particular problems. Mr. Franklin then discussed a number of designs typical of the examples in chemical technology.

STUDY SECTION OF THE ROYAL STATISTICAL SOCIETY

Reports of Meetings

LONDON

9th January, 1952. Population Projections: P. R. Cox

This was a joint meeting with the Students' Society of the Institute of Actuaries. The first part of his talk formed the basis of Mr. Cox's article in this issue of *Applied Statistics*. In the second part Mr. Cox was concerned mainly with the considerations governing the choice of alternative mortality and fertility assumptions for population projections. The different methods of projecting mortality rates into the future were illustrated with reference to the work of the Royal Commission on Population and of Australian and American demographers. Selecting bases for fertility was shown to be associated closely with the study of fertility and reproduction generally, and a brief survey was made of current problems in this field.

The substance of the remarks made in the second part of his paper has been incorporated in an article in the current issue of the *Journal of the Institute of Actuaries Students Society*.

13th February, 1952. Road Accident Statistics: F. Garwood

Dr. Garwood said that there were two ways of studying accident causation. One method was for the police to consider the facts about each accident and to assign a cause, or several causes, from a predetermined list. An alternative method, which the Road Research Laboratory favoured, was to compare the frequency of accidents under different conditions. Dr. Garwood suggested that while the first method was relatively easy to operate, it was based on a rather loosely defined concept: that of the cause of any particular accident; and the overall picture obtained from the statistics was liable to be misleading. To define the concept more precisely it would be necessary to ask if the accident would or would not have happened if each of a list of specified conditions were absent, and this was often impracticable.

On the other hand, the second method, though more reliable, had a rather more limited field of application, could not be guaranteed to give positive answers, and ignored a lot of useful facts or opinions which the police might record about each accident. In giving examples of the second method Dr. Garwood mentioned that before the war the accident rate had fallen in those police districts in which experimental motor patrols had been in operation. Also a comparison of the numbers of people walking on the two sides of the road where there were no footpaths with the numbers involved in accidents showed that it was safer to walk facing the traffic. As a third example he pointed out that on days in May when local elections were held the numbers of accidents to children increased considerably, presumably as a result of the schools being closed for polling stations.

12th March, 1952. The Census of Distribution: D. A. Clark

Mr. Clark described the Census and discussed the uses to which its results might be put.

BRISTOL

15th January, 1952. The Notion of Probability in Statistics: S. Körner

Dr. Körner reviewed various theories of probability, from the classical theory usually connected with Laplace to the frequency theories put forward by von Mises and Reichenbach. He believed that in spite of great progress in the solution of mathematical problems of probability the logical problem still awaited its solution, and suggested that those who hold the view that the probability calculus works and does not therefore stand in need of logical clarification are possibly closing the door on efforts to make it work even better.

19th February, 1952. The National Food Survey: W. L. Kendall

Mr. Kendall explained that the National Food Survey during the ten years of its existence had both extended its objects and developed its methods. Its original purpose was to check the nutritional level of diets in those households likely to be most affected by the wartime shortages, and it was limited to the urban working-classes and to gathering data coming within the direct knowledge of the housewife; meals outside the home and personal expenditure, for example on sweets, were excluded. The form of the survey was a weekly record collected from addresses selected at random in working-class wards in towns stratified according to region and population. Numerous recalls were made by the investigators to assist the housewife, and larder stock changes were measured in order to estimate the quantities of food consumed during the week.

Apart from the limitations already mentioned, defects in the resulting data arose from the large number of substitute addresses which the investigator found necessary to locate on the spot; from the bias in favour of the large household, which was the result of housewives in such households being more frequently at home when the investigator called; and from the distortion of the normal expenditure habits of the household brought about by the preoccupation of the housewife with the survey.

Beginning in 1950, the methods have been modified and the purpose of the survey has been widened as problems of the demand for food in general, and the special problems of marginal households, had become of increasing interest. Accordingly the sample has been extended to include the whole population but by adopting a number of simplifications and with a very much improved response rate (there were now no substitute addresses) it had been possible to provide a continuous monthly survey, as formerly, within the same resources. Further, the burden on the housewife had been eased by the introduction of a much simpler record of larder stock, with the result that the recorded conduct of the household was now more representative of usual food expenditure habits.

With the new and extended data, fresh problems of classification and analysis arose. Family composition was of special importance since this was found to be as powerful an influence as social class. The difficulty was to isolate large, homogeneous, and representative family composition groups,

and the present method was to examine chiefly the influence of children by classifying all households with one male and one female adult (about 60 per cent. of all households) by this variable.

Social class analysis at the present time raised other difficulties, since the levelling of incomes, the increase in the number of family earners, controlled prices, and the policy of subsidies, led to a narrowing and blurring of class distinction. The current method of classification was to group households by the incomes of the heads of households, but further information on total family incomes was required to provide a more satisfactory treatment. Food expenditure grouping was least satisfactory since it neglected both income and needs, and the results grouped in this way were difficult to interpret.

18th March, 1952. The Assumption of Randomness in Applied Statistics: G. F. Todd

After pointing out that the null hypothesis and exact significance tests, which are used in a large proportion of modern statistical work, fundamentally require that the data should be obtained by samples drawn at random from infinitely large populations, Mr. Todd considered the interpretation to be given to test results that proved to be statistically significant and the extent to which modern sampling procedures fulfilled the fundamental requirement.

Mr. Todd said that a statistically significant difference between observation and hypothesis could mean that one or more of three things had happened: that a very exceptional occurrence had in fact happened, that the data used were not random samples from an infinitely large population, or that the hypothesis was false. He suggested that insufficient attention had been given to the second possibility and said that the significance test really tested *two* hypotheses, namely the null hypothesis which had been explicitly formulated and the implicit hypothesis that the data were random samples. A significant difference indicated that one of these hypotheses was probably erroneous but did not say which.

Mr. Todd pointed out that there were four main ways, neglecting various refinements, in which the individual units of a population might be selected to form a sample. These were random sampling, systematic sampling, quota sampling, and bulk-selection sampling. In this last type of sampling, all or part of a limited sector of the population, selected neither at random nor systematically, was taken. An example was to take all the sixth-form boys in the two nearest schools as being a representative sample of sixth-form boys in all the schools in the country.

After considering random, systematic, and quota samples Mr. Todd went on to examine whether bulk-selected samples could be regarded as random samples if the method of selection was independent of the characteristic under consideration, and said 'It is therefore impossible to avoid the conclusion that in practice, no less than in theory, independence of the method of selection from the characteristic under consideration is not by itself sufficient to establish randomness in sampling.' Consequently, in view of the reliance commonly placed on this as a criterion of randomness, significance tests sometimes gave rise to what might be called 'nonsense differences' between hypothesis and observation, in the sense that significant differences had been disclosed in circumstances where the significance tests were strictly inapplicable.

BOOK REVIEWS

W. J. Youden: *Statistical Methods for Chemists.* New York: John Wiley & Sons; London: Chapman & Hall, 1951. Pp. x + 126. 24s.

This book represents an attempt to give the analytical chemist of some maturity an idea of the application to his work of significance testing and experimental design. The style is discursive rather than systematic; formulae are very infrequent, and techniques are presented mainly through examples.

The author soon plunges into a treatment of small sample techniques and shows how to calculate and compare standard deviations. Then come the comparison of averages by the *t*-test, the *F*-test, two-stage sampling, the fitting and comparison of regression lines, and the use of two-factor analysis of variance. Chapter 7 explains systematic and random interactions, showing how they arise, and ends with a schematic presentation of the analysis of variance for three factors with replications, with remarks on its interpretation. Chapter 8 states the case for planning experiments beforehand and illustrates some properties of randomisation, showing how it can justify the use of the *t*-test. The reader is then introduced to various types of experimental design—a subject to which the author has contributed. Although there are very few diagrams there are twenty real-life examples which should strike a sympathetic chord in chemists. There are many remarks about the relevance of the techniques to general situations in chemists' experience.

In the reviewer's opinion, the book will be most useful when studied with a standard introductory text-book, when it may help chemists a good deal in applying statistics. It does not speak sufficiently clearly about principles to stand alone as an introduction, nor is it comprehensive or systematic enough to serve as a general reference book. G. H. JOWETT

L. H. C. Tippett: *Technological Applications of Statistics.* London: Williams & Norgate; New York: John Wiley & Sons, 1952. Pp. vii + 189. 18s.

This book, although of limited scope, is well balanced. The first part is concerned mainly with the basic theory of control charts, the details of the procedure for their application, and acceptance sampling. The second part commences with a discussion of significance testing, the precision of an average, and the use of confidence limits. The application of the analysis of variance and its relation to experimentation are considered and there is a short chapter on correlation and regression. In the planning of investigations Mr. Tippett considers randomised blocks and Latin squares. The chapters on analysis of variance and planning of experiments are rather limited in scope, but they are sufficient for most of the simpler investigations.

In a book of this sort it is important that the basic principles are fully explained, and this is one of the strong points of Mr. Tippett's book. Particularly well presented are the chapters on significance testing and the precision of estimates and confidence limits, although the reviewer would have liked a greater shift of emphasis to the use of confidence limits at the expense of tests of significance.

Mr. Tippett intends this book as an introduction and companion to a systematic text-book or course of study in applied statistics. In this he has been successful. The book is well suited for a first reading of statistical methods for industrial technicians and students. O. L. DAVIES

THE SCOPE OF MEDICAL STATISTICS

NORMAN T. J. BAILEY
Department of Human Ecology, University of Cambridge

In surveying the applications of statistics to medicine, Mr. Bailey deals with subjects of general interest: the use of medical records, the making of efficient arrangements for providing hospital services, and the testing of the efficacy of medical treatments, as well as with the specialist subjects of epidemiology, the efficiency of techniques of medical investigation, and medical genetics. The article is based on a paper given at a conference of the Royal Statistical Society held in Cambridge in September 1951.

Introduction

Medical statistics may, I think, fairly be said to have started in the second half of the seventeenth century with the work of John Graunt and William Petty, who based many of their observations and conclusions on the famous London Bills of Mortality. Although vital statistics, with suitable modern refinements, is still an important part of the whole corpus of medical statistics the latter has, in the last 200 years, gradually increased its scope. With the assimilation of the advances of modern mathematical statistics, especially those occurring in the last thirty years, medical statistics may now reasonably claim applications to the whole realm of medical science.

It is of course impossible in the space available to review more than a small proportion of the whole field, but I shall try to give as wide a selection of topics falling within the province of medical statistics as possible. My object is twofold: firstly, to illustrate the variety and scope of the subject; and, secondly, to indicate a number of problems which seem worthy of further study and understanding. I make no special apology for drawing on much material with which I have been personally concerned—partly because it is easier to do justice to subjects to which one has given particular attention, and partly in order that my remarks should reflect the statistical interests of the Cambridge Medical School.

Medical Records and Socio-medical Surveys

A place of fundamental importance in medical statistics is held by that part of the subject which is based on medical records. Much valuable information can be gleaned from a consideration of mortality and morbidity rates for various diseases, especially in relation to environment and social class. An enormous amount of work of this kind is regularly carried out at various centres and it would be impossible to make an adequate summary. There are, however, a number

of useful introductions to the potentialities of this approach.[1, 2, 3] Reliable data are of course nearly always essential in any kind of investigation, but when, in contradistinction to comparatively simple laboratory readings and measurements, one is concerned with the analysis of vast amounts of information collected at considerable expense over long periods of time about thousands of individuals, then some special attention to methods of recording and analysing the data is called for.

One of the most obvious examples is the collection of hospital statistics, which are particularly useful, not only for the routine collection of standard morbidity and mortality data, but also in the study of case-loads for administrative purposes to see, for example, whether the supply of medical facilities is adequate for meeting the demand. Again, a hospital should have an efficient records system so that the notes relating to a given patient can be easily found, or so that all the cases of a given disease can be extracted for detailed study. On the other hand, the use of an elaborate punched-card system, with the object of doing detailed retrospective research into the effects of various medical and surgical treatments, is in my opinion rarely justifiable. In the first place, the valid comparison of different methods and treatments usually requires carefully controlled experimental conditions, which are hardly ever found to have occurred naturally in a retrospective survey. Secondly, it is extremely difficult, if not impossible, to design a record form which will cover adequately all the data which may be needed from every kind of patient for some future research. It seems to me much more satisfactory to use a relatively straightforward and inexpensive system for producing routine statistics. Then if a particular piece of research is to be undertaken, special record forms can be designed and all the paraphernalia of punched-card equipment brought into action.

The kind of investigation for which punched-card methods are invaluable is the socio-medical type of work where a great deal of clinical, social, and economic data is collected about each of a large number of patients suffering from a particular disease, together with a number of controls, living in a broadly specified environment. Extensive record forms can be specially designed to serve as both original clinical records and transcription forms for punched-card work. It should be said that with the most modest range of equipment, say a hand-punch and a counter-sorter only, quite extensive surveys of this type can be undertaken and conveniently analysed if one makes a special effort to compress all the data for a single individual onto one punched card. When several cards are used for each individual the correlation of data on different cards usually entails special machinery which greatly adds to the cost of the simple set-up referred to above.

The Provision of Treatment

Under the general heading of 'provision of treatment' come all the

problems connected with efficiency of supplying patients or prospective patients with known methods of remedial or preventive treatment. We want to know whether the general practitioners, specialists, nurses, social welfare workers, and so on, are available in sufficient numbers for the needs of the particular community they work in. We also want to know whether the services in which these individuals play a vital role are efficiently organised. This is a particularly pressing problem with present-day shortages of trained staff and housing accommodation. Very little building of new hospitals, clinics, and health centres can be done in the near future, and the main opportunities for the better provision of medical treatment, apart from fundamental advances in actual medical knowledge, will probably be found in the more efficient utilisation of existing resources.

The question of hospital function and design, together with various allied problems, is the object of a special study which is being made by a research team under the auspices of the Nuffield Provincial Hospitals Trust.[4]

One particularly important topic is the rationalisation of traffic flow within a hospital. This is concerned with reducing to a minimum the movement from one part of the hospital to another of patients, nurses, doctors, visitors, and equipment, and is desirable on grounds of both convenience and hygiene. One can conceive a hospital as consisting of a number of units. There are the beds, duty rooms, utility rooms, lavatories, etc., for each ward-unit in several specialties. In addition there are a variety of special services and facilities: operating theatres, kitchens, X-ray department, pathology laboratory, and so on. On medical grounds certain types of traffic from one place to another will be essential: we can form a general pattern of the necessary traffic in terms of the number of journeys that have to be made from one point to another in a certain period of time. At its simplest the problem is: how to site all the various units so as to reduce to a minimum the total volume of traffic—measured in units involving distance as well as numbers of persons or amounts of material. There are several complications, of course, including the difficulty of coping with peak-hour requirements. A general solution of the simplest problem, which would involve a combination of statistics and the calculus of variations, is not yet available. An *ad hoc* procedure is to draw up a number of alternative designs which are acceptable both architecturally and medically, and then to apply to each the standard pattern of traffic with the object of discovering which design involves the smallest total flow. A fair degree of success has already been obtained with this approach, both with regard to the relative siting of some of the major units within a hospital, and also with regard to the arrangement within a ward-unit so as to cut down the total walking distances of nurses on duty.

Another problem is concerned with deciding how many beds in a ward-unit should be separately accommodated in single rooms. Separate accommodation is desirable on medical grounds for dying

patients, those who require special treatment, patients who are highly infectious or who are peculiarly susceptible to infection, and so on. Dr. Goodall[5] has discussed this problem and has made a further distinction between separately accommodated patients who must be immediately accessible and those who need not be. The number of single rooms available is of course a fixed quantity, but the demand shows considerable day-to-day fluctuations. It is evident that a simple estimate of the average requirement will not give a proper indication of the number of single rooms that should be provided and I have undertaken a statistical treatment.[6] For any given number of single rooms provided we can calculate (1) the extent to which the demand for single rooms can be met—the *efficiency of provision*—and (2) the extent to which the single rooms actually provided would be occupied for the proper purpose—the *efficiency of utilisation*. The former can be made very high only at the expense of the latter, and vice versa; for if enough single rooms are provided to cope with nearly all requirements even on 'peak' days, many will not be used for their proper purpose on other days. If a satisfactory balance is to be struck between these conflicting factors, namely the purely medical needs and the necessity for economising space and finance, then hospital planners must try to choose an optimum number of single rooms, giving as high an efficiency of provision as possible without allowing the efficiency of utilisation to drop too low. I found, for example, that with an average 16-bed surgical unit two single rooms should be provided for patients who must be easily accessible: the two rooms would satisfy 84 per cent. of the demand and be used 73 per cent. of the time; three rooms would satisfy 92 per cent. of the demand but would be used for their proper purpose only 59 per cent. of the time, and so on.

Most people will immediately regard the design of appointment systems for out-patient clinics as being of some urgency. The number of people waiting for medical examinations and the rate of examination both bear very strongly on the design of the department. It is important to have a good appointment system, not only to reduce waiting times to a minimum, but also to use the available accommodation in the best possible way. It is not uncommon for the average consultation time to be of the order of ten minutes while the average waiting time is upwards of an hour.

In the queueing problems so ably discussed by Mr. D. G. Kendall[7] the main attention was concentrated on the situation where customers arrived at random. In out-patient clinics, on the other hand, the patients can, within limits, be made to arrive at predetermined times, while the consultation time is variable, having some characteristic frequency distribution.

The basic idea is that, given the distribution of consultation time, we ask what is the optimum appointment system which will save the patients the greatest amount of waiting time without the consultant having too long an idle period. Theoretical work of a fairly general

nature by Mr. Lindley[8] has made available the limiting distribution of waiting times for an infinitely long queueing process when the average consultation time is less than the appointment interval. Using the alternative approach of random number studies I[9] have investigated the case of relatively short queueing processes, involving say 25 patients. Assuming a suitably chosen Pearson Type III curve for the frequency distribution of consultation times I showed that an optimum appointment system is as follows. Suppose that the average consultation time is 5 minutes. Then patients are given appointments at regular 5 minute intervals, the consultant commencing work as the second patient arrives. For a clinic of 25 patients the patients' average waiting time would be 9·1 minutes, while the consultant would waste, on average, 5·7 minutes per clinic. A further discussion of the practical applications has been made by Brigadier Welch and me.[10]

The Efficacy of Treatment

As the primary object of medical science is either to treat or prevent disease and disablement, so one of the main applications of medical statistics is in the testing of such measures. Actually, as I see it, the real difficulty is not a statistical one at all, but consists in the widespread need for arranging that adequate tests be done. Although, of course, many drugs in current use have been subjected to properly controlled tests, it is equally true that large numbers are, at least statistically speaking, of doubtful or unknown value. It is not easy to organise genuine experiments on human subjects, and in any case the responsibility for initiating such work does not lie with the statistician. However, the statistician can do much to improve the situation by encouraging those whose main concern is with the discovery and application of new methods both to examine more critically existing medical procedures and not to accept new ones without first carrying out stringent tests.

Given the opportunity to undertake such tests it is usually not difficult to apply standard statistical methods to the design and analysis of the appropriate medical data, whether it is estimating the recovery or relapse rate for a new treatment or comparing the degrees of protection against certain diseases given by various vaccines, and so on. If one has sufficient patients—who may have to be volunteers—then it is often possible to run the standard 'factorial experiments' involving several treatments on different types of patient under a variety of environmental conditions. However, it is a salient characteristic of medical work that it is often not easy to get a sufficient number of individuals to obtain conclusive results. The disease in question may be uncommon, and volunteers may be few and far between. This has led to the increasing use of 'matched samples,' whereby several different treatments are tried on each individual. With quantitative data this can be taken into account by regarding the individual as a block of plots, as in agricultural experiments. But with qualitative data special

methods are required. With appropriate safeguards the method of matching appears to lead to an increase in the precision of the results obtained. For example, J. E. Denton and H. K. Beecher[11] in their tests of a new analgesic drug gave the drug and a control injection at different times to every patient. The mathematical implications of matched samples do not yet seem to have been fully explored, although a valuable discussion of significance tests in the multiple-sample case has been given by Professor Cochran,[12] who has extended the two-sample test originally given by Q. McNemar.[13]

Another statistical technique which does not yet seem to have been made full use of in medical research is *sequential analysis*.[14] As is well known, this method involves sampling according to a prearranged plan, whereby one stops the sampling process as soon as a definite decision is obtained. The process continues to be applied only so long as the criteria adopted still indicate that no decision is yet possible. In general, sequential methods require smaller samples to achieve a given result, so that the need for such methods in medical work is immediately obvious. Another advantage, which so far as I am aware has not been completely realised, is that the use of a sequential method would afford some relief from the perennial question: 'When should a promising new drug, which is not yet fully tested, be given to every patient?' This is a very real ethical complication which adds to the other difficulties of carrying out satisfactory tests of medical treatments and I do not suggest that the use of a sequential test would avoid all the moral pitfalls; but it would enable the criteria for decisions to be discussed and agreed upon before the experiment started—that is the important point. It would avoid adding to the doctor's burden of responsibility by making him try to decide during the course of an ordinary non-sequential test just whether he ought to continue or break off the experiment at any stage. We might, for example, be comparing the recovery rate in batches of hospital patients, half being given the old and half the new drug, and could decide before the experiment precisely what action would be taken if any particular series of recovery rates were observed for the two drugs. A sequential procedure could then be drawn up which satisfied both medical ethics and the statistical requirements. This would ensure that whatever the outcome of the experiment at any particular stage, the decision taken would be the best possible.

Epidemiology

The existence of widespread infections or contagious diseases has always been regarded as one of the major scourges of mankind and, judging from their prevalence in both animal populations and primitive human communities, such infections and diseases have presumably been with us since our earliest origins. Many factors contribute to modify the behaviour of epidemic outbreaks in the course of time—old diseases may become less virulent or show changes in age incidence and

new diseases of obscure origin may arise. The operation of natural selection in eliminating the unduly susceptible; improvements in general health, nutrition, and hygiene; advances in methods of treatment; and so on; all these make their contribution to the fight against epidemic disease. In spite of great improvements the problem is still a serious one. Although the general picture is for many diseases reasonably clear, detailed quantitative knowledge about such factors as the actual mechanism of infection, transit times, incubation times, variations in infectivity, numbers of susceptibles, and carriers in the population are often very obscure.

Epidemiologists have always made great use of the so-called 'epidemic curves'—that is curves giving the numbers of new cases of disease occurring each day or week in a given area—in the hope that they will reveal facts about the character of the epidemic that would otherwise not have been obvious. There does seem to be a reasonable hope of advance in this direction, but so far many of the conclusions drawn are still only speculative. The basic principle is that we construct various theoretical epidemic curves based on a variety of different assumptions and then see which fits actual data most satisfactorily.

Most of the theoretical work done in the past has been of a 'deterministic' nature, that is, it has been assumed that, for given numbers of susceptible and infectious individuals and given infection, death, and recovery rates, a certain definite number of fresh cases would arise in a given time. The early work of Ross[15–17] on what he called '*a priori* pathometry' and of Brownlee[18] was of this type. So also were the much more elaborate studies later undertaken by Kermack & McKendrick[19] and Soper.[20] Kermack & McKendrick showed that for a given set of infection and recovery rates there was a certain threshold density of population. If the actual population density were initially below this value the introduction of an infected person would produce no epidemic; if the density were above the threshold an epidemic would occur in such a way as to reduce the density of susceptibles as far below the threshold as originally it was above. With diseases transmitted through a host there is a similar threshold theorem involving the product of the densities of the populations of man and host.

Soper's work was mainly on the periodicity observed in the recurrence of epidemics in diseases like measles—again some insight was gained into the underlying mechanism and it was evident that the continual entry into schools of young susceptibles was one of the important factors.

Little success has in the past been obtained by trying to predict the course of an epidemic by fitting curves to the initial data. In any case curve-fitting can hardly be expected to afford any clue to the physical and biological processes taking place unless it is based on some particular sort of mathematical model.

One of the difficulties inherent in the notified returns of a disease is that they usually relate to such large areas. It is well known that an

overall epidemic can often be broken down into smaller epidemics occurring in separate regional subdivisions. These regional epidemics are not necessarily in phase and may interact with each other. Taking the process of subdivision a stage farther we can consider a single borough or district, or finally, the 'effective group' in which a single individual moves, since in many cases he will be in close contact with a small number of people only, say about 10–50. The observed epidemic for the whole region can then be thought of as built up from a number of small epidemics taking place simultaneously in several relatively small groups of associates and acquaintances. Although in practice the groups may overlap, we can employ the concept of an 'effective' number of independent groups or isolates. A typical simplified model involves a community of k independent groups each of size n. We imagine an epidemic started by the simultaneous appearance or introduction of k primary cases, one for each group. Now, for comparatively small values of n, statistical fluctuations will be appreciable; the older deterministic treatments will be unsuitable and we shall have to appeal to so-called 'stochastic' methods. Professor Bartlett[21] has emphasised this need and has devoted some discussion to various partial attacks already made on the problem. He[22] has also referred to the simplest stochastic epidemic in which only infection but not recovery is considered, and I have discussed this problem in greater detail.[23] It should be mentioned that these results for epidemics with no recovery are in fact also applicable to epidemics for which the time taken to recover is fairly long compared with the period occupied by the humped part of the epidemic curve. With the assumptions made previously about epidemics taking place simultaneously in separate groups, it is evident that with a fairly large number of groups the epidemic curve for the whole region will resemble in shape the curve corresponding to the stochastic mean for a single small group. In epidemic processes, it should be remarked, stochastic means are not the same as the corresponding deterministic values.

The mathematical treatment of epidemic processes presents many analytical difficulties not the least of which is to find expressions, for the stochastic mean, for example, which are suitable for computation and practical application. It may be that the greater chance of progress lies in the alternative approach of using random number studies.

A word should be said, in passing, about the possibility of experimental work on epidemics, using a suitable animal such as the mouse. Work of this sort has been very skilfully undertaken by Greenwood, Topley & Wilson and has included studies of the waxing and waning of epidemics in infected herds, the effect of introducing new unprotected immigrants, and the effect of withdrawing animals from the herd.[24] Mouse communities exhibit many of the phenomena observed in human communities with regard to the occurrence of epidemic disease. The advantage of the animal work is that the time scale is short and conditions can be closely controlled. There would seem to be great

possibilities in further work of this kind for increasing our understanding of both human and animal epidemics. Incidentally, properly designed experiments coupled with appropriate statistical analysis might shed some fresh light on the role of vaccination in the prevention of disease—e.g. smallpox vaccination in man and fowl-pest vaccination in fowls.

Another way in which statistical analysis can be of use to the epidemiologist is in the study of the distribution of multiple cases of disease in households. It is well known that measles tends to show an approximately biennial periodicity. In the discussions of this by Hamer and later by Soper it was shown how such a periodicity could arise from a number of assumptions, one of which was that a given case was highly infectious for a short period of time only. Can some test be made of this latter hypothesis ?

Broadly speaking we want in the first instance to differentiate between the situation in which we expect a straightforward binomial distribution of secondary cases, e.g. groups of people drinking typhoid-infected water, and the situation in which the disease is passed from person to person. Discussing a measles epidemic occurring in St. Pancras in 1926, Greenwood[25] showed that the first hypothesis of a binomial distribution of cases was quite inadequate. Now consider the second hypothesis. The argument used by Greenwood was that the distribution of secondary cases resulting immediately from the first case would be binomial; but then susceptibles escaping this risk might be infected later by the secondary cases themselves, giving rise to tertiary cases, and so on. This is the principle of the so-called 'chain' binomial argument. It should be valid when a disease is highly infectious for short periods. Greenwood estimated the chance of infection by fitting mean values. How efficient a procedure this is seems never to have been investigated, though it was evident that quite good agreement was obtained. A more sophisticated test of significance for seeing whether the multiple cases in a household are distributed at random should be derivable from a discussion of the highly interesting mathematical problems involved in the random division of an interval.[26, 27]

Further problems of the same type arise from asking whether the geographical distribution of a disease is random—compare, for example, the discussion by Dr. Cruickshank[28, 29] of the incidence of tuberculosis and cancer in the UK, and the statistical treatment of so-called 'statistical maps' by Professor Moran.[30]

I believe that much important work remains to be done in connection with both the household and geographical distribution of cases. These problems certainly present several theoretical difficulties, especially in regard to finding convenient tests of significance and efficient methods of estimation, though these difficulties are no doubt not insuperable.

Increasing the Efficiency of Medical Techniques

Basically, one considers whether the technique employed achieves

the greatest accuracy to be expected when the inevitable statistical fluctuations are taken into account. If this optimum has not been reached it may be possible to suggest improvements that will enable better results to be obtained. The use of a modified technique with greater intrinsic accuracy means either that one can obtain more accurate estimates for the same effort, or that one can achieve the same overall accuracy as before with less effort.

Let us consider as a typical example the routine examination of differential white cell counts. With good technique we should expect a multinomial distribution of the proportions of the different kinds of leucocyte. In practice the observed accuracy of the percentages of different kinds of cells, polymorphonuclear cells, lymphocytes, monocytes, etc., may be less than that expected, owing to various disturbing factors. These may be the tendency of the cells to clump or the liability of certain types of cells to be displaced to particular portions of the slide. Indeed, the latter phenomenon will lead to systematic errors as well unless special steps are taken to counteract it, such as the employment of the so-called 'battlement' count. Moreover, it may be found that with certain techniques the expected accuracies are achieved only if the total number of cells counted exceeds a certain level. For example, in discussing a new mechanical aid for making blood films Marks, Bailey & Gunz[31] showed that for the new method the estimate of the proportion of polymorphonuclear cells was effectively binomial in distribution for counts as low as 100 (actually unpublished data show this may be extended down to as low as 50). With the old method of spreading the films, on the other hand, the binomial distribution held only if the total count was 300 or more.

Again, the problem may be not so much to improve the technique as to find a more efficient method of analysing the data. A much-discussed example of this occurs in the field of biometrical assay where one is assaying the potency of a drug by considering the series of mortalities obtained in batches of test animals for different strengths of the drug concerned. In large samples at any rate, with a normal distribution of tolerances, the greatest efficiency in assessing the data is obtained by using the methods of probit analysis.[32] For a single series of experiments designed to answer certain specific questions it is undoubtedly worth the effort of undertaking the full-dress statistical analysis, but very often it is desirable to have approximate estimates which can be found quickly in the course of preliminary laboratory experiments. Several such alternative methods have been suggested and widely used in the present context. A good deal of work has been done on the relative merits of these alternative procedures.[33, 34]

Medical Genetics

Although genetics stands on its own as an autonomous field of study, certain results, particularly those concerned with human heredity, are of considerable importance to medicine. For the adequate assessment

and interpretation of a great deal of genetical data a fair amount of mathematics is often essential, and thus it comes about quite naturally that medical statistics finds itself involved in genetical problems.

First, there are those *population* studies carried out mainly in the late '20's and early '30's by men like Sir Ronald Fisher, Professor Haldane, and Professor Wright. Evolutionary progress was to be explained by the action of natural selection on the genetically diverse material provided by random mutation. Specific gene models were considered and for given mutation-rates and selection pressures the evolutionary consequences could be ascertained by mathematical analysis. This approach can be used in considering the long-term evolutionary changes in the general level of health or disease in the population as a whole. One application was a discussion by Professor Haldane[35] of the selection against heterozygotes owing to the occurrence of haemolytic disease in connection with the Rhesus blood-group. Another application is in the estimation of the harmful effects which may appear in future generations if sources of irradiation like X-rays and atomic energy are used irresponsibly.[36, 37]

Secondly, there are what we may call *individual* studies. These deal with the inheritance of common well-defined characters like blood-groups as well as with inherited diseases and abnormalities. Doctors are frequently consulted by patients about the possibility of their developing or handing on to their children some disease or defect which is known to be in the family. Popular ideas in this matter are often quite unnecessarily alarmist. For example, suppose a man who wishes to marry his first cousin had a brother who died from juvenile amaurotic idiocy. The man and his cousin are both healthy and want to know the risks with respect to this disease if they marry and have children. Juvenile amaurotic idiocy is inherited as a simple autosomal recessive, and it can be shown by examining the pedigree in this particular case that there is 1 chance in 6 that both the man and his cousin are carrying a 'bad' gene. Thus the chance of the first child being affected is 1 in 24. The couple may well be prepared to take this chance, but it is desirable for them to know what it is. More complicated situations may require more careful consideration of the probabilities involved. This sort of eugenic prognosis could be made very much more precise if common characters were known to be genetically linked to the abnormality in question. A good example is available involving the not infrequent character of colour-blindness. It is usually not possible to say more than that the childless daughters of a haemophilic man have a 1 in 2 chance of carrying the haemophilia gene and passing it on to half their sons. Now the loci for haemophilia and colour-blindness are linked. In a certain Dutch family, where both these abnormalities were liable to appear either separately or together in certain individuals, it was possible to estimate that the colour-blind daughter had 9 chances in 10 of carrying a haemophilia gene while her normal sisters had only 1 in 10. It is clear that the discovery of such linkages is of great

importance, but, apart from sex-linkages, few if any have yet been definitely established in man.

Not only are human families relatively small but also it is difficult to find marriages between individuals of the appropriate genetic constitution: even then we may know the mating-type only with a certain degree of probability. It has, however, proved possible to develop powerful statistical methods for examining material of the right kind for the existence of genetic linkage. These methods vary according to whether the data consist of a number of isolated pedigrees each containing several individuals related by various degrees of kinship, some of whom may be affected with the abnormality in question, or whether we have a series of specially selected compact family groups, each normally consisting of just parents and children. For the former type of material the methods used by Professor Haldane[38] are appropriate: a particularly useful discussion is the one by Haldane & Smith[39] on estimating the linkage of haemophilia and colour-blindness. For the second type of material Professor Penrose's sib-pair method[40, 41] has been widely used, though it is in general less efficient than the method of u-statistics.

The latter method was first developed by Sir Ronald Fisher[42–44] and has the advantage of being efficient and enabling data from families of varying size to be combined directly. Dr. Finney[45, 46] subsequently extended the method to cover a wide variety of mating types and to deal with such complications as multiple allelomorphs and incomplete parental records. He provided conveniently tabulated lists of the appropriate scores and information functions. I have made a further extension of the method to the cases of rare abnormalities where the gene or genes, though present, may fail to be expressed in a certain proportion of individuals[47, 48]—as with Huntington's chorea for example. Another important problem is the estimation of population gene frequencies. When the individuals in a sample are related, as in data collected from family records, special methods of analysis are required.[49–52]

A great deal of the work in medical genetics is concerned with the inheritance of common characters like the nine main blood-group systems and the ability to taste phenylthiocarbamide. Knowledge of the modes of inheritance of blood-groups and also the corresponding gene frequencies are valuable not only for questions like estimating the probability of the risk of haemolytic disease in babies of Rhesus negative women with Rhesus positive husbands, but also in medico-legal work. The main applications of the latter are to problems of relationship, especially those involved in disputed maternity or paternity.

REFERENCES

Medical Records and Socio-medical Surveys

[1] Ryle, J. A. (1948). *Changing Disciplines.* Oxford University Press.
[2] Cluver, E. H. (1952). *Social Medicine.* H. K. Lewis, London.
[3] Medical Research Council (1951). *The Application of Scientific Methods to Industrial and Service Medicine.* HMSO, London.

The Provision of Treatment

[4] Nuffield Provincial Hospitals Trust (1951). Report 1948–51.
[5] Goodall, J. W. D. (1951). 'Single Rooms in Hospital.' *Lancet*, **1**, 1063.
[6] Bailey, Norman T. J. (1951). 'On Assessing the Efficiency of Single-room Provision in Hospital Wards.' *J. Hyg. Camb.*, **49**, 452.
[7] Kendall, D. G. (1951). 'Some Problems in the Theory of Queues.' *J. R. Statist. Soc.*, **B**, **13**, 151.
[8] Lindley, D. V. (1952). 'The Theory of Queues with a Single Server.' *Proc. Camb. Phil. Soc.*, **48**, 277.
[9] Bailey, N. J. T. (1952). 'A Study of Queues and Appointment Systems in Hospital Out-patient Departments, with special reference to waiting-times.' *J. R. Statist. Soc.*, **B**, **14** (in the press).
[10] Welch, J. D., and Bailey, N. T. J. (1952). 'Appointment Systems in Hospital Out-patient Departments.' *Lancet*, **1**, 1105.

The Efficacy of Treatment

[11] Denton, J. E., and Beecher, H. K. (1949). 'New Analgesics.' *J. Amer. Med. Ass.*, **141**, 1051.
[12] Cochran, W. G. (1950). 'The Comparison of Percentages in Matched Samples.' *Biometrika*, **37**, 256.
[13] McNemar, Q. (1949). *Psychological Statistics.* John Wiley and Sons, New York.
[14] Wald, A. (1947). *Sequential Analysis.* John Wiley and Sons, New York.

Epidemiology

[15] Ross, R. (1916). 'An Application of the Theory of Probabilities to the Study of *a priori* Pathometry, I.' *Proc. Roy. Soc.*, **A**, **92**, 204.
[16] *Idem* (1917). 'An Application of the Theory of Probabilities to the Study of *a priori* Pathometry, II.' *Ibid.*, **A**, **93**, 212.
[17] Ross, R., and Hudson, H. P. (1917). 'An Application of the Theory of Probabilities to the Study of *a priori* Pathometry, III.' *Ibid.*, **A**, **93**, 225.
[18] Brownlee, J. (1918). 'Certain Aspects of the Theory of Epidemiology in Special Relation to Plague.' *Proc. Roy. Soc. Med.* (Sect. Epid. and State Med.), 85.
[19] Kermack, W. O., and McKendrick, A. G. (1927 and later). 'Contributions to the Mathematical Theory of Epidemics.' *Proc. Roy. Soc.*, **A**, **115**, 700; **138**, 55; **141**, 94.
[20] Soper, H. E. (1929). 'Interpretation of Periodicity in Disease-prevalence.' *J. R. Statist. Soc.*, **92**, 34.
[21] Bartlett, M. S. (1949). 'Some Evolutionary Stochastic Processes.' *Ibid.*, **B**, **11**, 211.
[22] *Idem* (1946). *Stochastic Processes* (notes of a course given at the University of North Carolina, 1946).
[23] Bailey, Norman T. J. (1950). 'A Simple Stochastic Epidemic.' *Biometrika*, **37**, 193.
[24] Greenwood, M., Hill, A. Bradford, Topley, W. W. C., and Wilson, J. (1936). *Experimental Epidemiology.* SRS No. 209 of MRC. HMSO, London.
[25] Greenwood, M. (1931). 'On the Statistical Measure of Infectiousness.' *J. Hyg. Camb.*, **31**, 336.

[26] Moran, P. A. P. (1947). 'The Random Division of an Interval.' *Suppl. J. R. Statist. Soc.*, **9**, 92.
[27] *Idem* (1951). 'The Random Division of an Interval, II.' *Ibid.*, **B**, **8**, 147.
[28] Cruickshank, D. Barron (1940). *Papworth Research Bulletin*, 36.
[29] *Idem* (1947). 'Regional Influences in Cancer.' *Brit. J. Cancer*, **1**, 109.
[30] Moran, P. A. P. (1948). 'The Interpretation of Statistical Maps.' *J. R. Statist. Soc.*, **B**, **9**, 243.

Increasing the Efficiency of Medical Techniques

[31] Marks, J., Bailey, N. T. J., and Gunz, F. W. (1950). 'A Mechanical Aid in Making Blood Films.' *J. Clin. Path.*, **3**, 168.
[32] Finney, D. J. (1952). *Probit Analysis* (2nd edition). Cambridge University Press.
[33] *Idem* (1950). 'The Estimation of the Mean of a Normal Tolerance Distribution.' *Sankhyā*, **10**, 341.
[34] Armitage, P., and Allen, Irene (1950). 'Methods of Estimating the LD50 in Quantal Response Data.' *J. Hyg. Camb.*, **48**, 298.

Medical Genetics

[35] Haldane, J. B. S. (1942). 'Selection against Heterozygosis in Man.' *Ann. Eugen. Lond.*, **11**, 333.
[36] *Idem* (1947). 'The Dysgenic Effects of Induced Recessive Mutations.' *Ibid.*, **14**, 35.
[37] Muller, H. J. (1950). 'Our Load of Mutations.' *Am. J. Human Genetics*, **2**, 111.
[38] Haldane, J. B. S. (1941). *New Paths in Genetics*. Allen and Unwin, London.
[39] Haldane, J. B. S., and Smith, C. A. B. (1947). 'A New Estimate of the Linkage between the Genes for Colour-blindness and Haemophilia in Man.' *Ann. Eugen. Lond.*, **14**, 10.
[40] Penrose, L. S. (1935). 'The Detection of Autosomal Linkage in Data which consist of Pairs of Brothers and Sisters of Unspecified Parentage.' *Ibid.*, **6**, 133.
[41] *Idem* (1946). 'A Further Note on the Sib-pair Linkage Method.' *Ibid.*, **13**, 25.
[42] Fisher, R. A. (1935). 'The Detection of Linkage with "Dominant" Abnormalities.' *Ibid.*, **6**, 187.
[43] *Idem* (1935). 'The Detection of Linkage with Recessive Abnormalities.' *Ibid.*, **6**, 339.
[44] *Idem* (1936). 'Tests of Significance applied to Haldane's Data on Partial Sex Linkage.' *Ibid.*, **7**, 87.
[45] Finney, D. J. (1940 and later). 'The Detection of Linkage.' *Ibid.*, **10**, 171; **11**, 10, 115, 224, 233; **12**, 31.
[46] *Idem* (1942). 'The Detection of Linkage.' *J. Hered.*, **33**, 157.
[47] Bailey, Norman T. J. (1950). 'The Influence of Partial Manifestation on the Detection of Linkage.' *Heredity*, **4**, 327.
[48] *Idem* (1951). 'The Detection of Linkage for Partially Manifesting Rare Dominant and Recessive Abnormalities in Man.' *Ann. Eugen. Lond.*, **16**, 33.
[49] Fisher, R. A. (1940). 'The Estimation of the Proportion of Recessives from Tests carried out on a Sample not wholly Unrelated.' *Ibid.*, **10**, 160.
[50] Cotterman, C. W. (1947). 'A Weighting System for the Estimation of Gene Frequencies from Family Records.' *Cont. Lab. Vert. Biol.*, Univ. Michigan, 33.
[51] Finney, D. J. (1948). 'The Estimation of Gene Frequency from Family Records. I.' *Heredity*, **2**, 199.
[52] *Idem* (1948). 'The Estimation of Gene Frequency from Family Records. II.' *Ibid.*, **2**, 369.

A STUDY OF THE LIVES OF POTS USED IN MAKING GLASS

JOAN KEEN and DENYS J. PAGE*

Communication from the Staff of the Research Laboratories of The General Electric Co. Ltd., Wembley, England

It was required to learn something of the factors that influence the life of clay pots used in making glass, and only the statistically 'unbalanced' records of two years' actual working were available as data. Miss Keen and Mr. Page describe how they solved the problem by repeatedly applying a simple method of analysis.

The Manufacturing Process

The melting of glass in quantities suitable for the mouth-blowing of certain lamp bulbs and other components is carried out in clay pots of the type shown in Plate I. These are inserted into a hot furnace by unbricking the furnace wall and are heated until the glass has reached the correct consistency for blowing or pressing.

The manufacture of these refractory pots is a skilled craft of which few exponents remain today, each craftsman having his own peculiarities which may give rise to considerable differences in pot performance. Each time a pot fails it is necessary to remove it from the furnace and replace it with a new pot preheated in a separate kiln nearly to furnace temperature and this results in undesirable cooling of the other pots in the furnace; it is therefore vital that lives should be as long and as uniform as possible in order to reduce such disturbances to a minimum.

It is impossible to do a planned experiment to provide data amenable to the analysis of variance, so we were given the task of analysing the data from two years' pot performance and suggesting how possible future changes in level consequent on technical improvements could best be detected and recorded.

The Manufacturing Variables

The formal analysis concentrated on only seven of the many variables present, superficial examination of the others showing that little improvement could be gained by incorporating them. For ease of presentation only a few of these variables and 'levels' within the variables are considered in this account, and Table I shows which these are. All these factors, singly or in combination, are able to influence the effective life of the pots to a considerable extent. It will be shown later how a pot with a particularly fortunate combination of

* Now with the Southern Electricity Board.

factors can last up to twenty times as long as one with the worst possible background.

TABLE I

Variables	No. of Levels in Analysis	Levels in this Account
Maker	6	Smith, Jones, Robinson
Furnace	6	A, B, C, D
Gas	3	Producer, town's
Size—type	8	Open, covered, half
Position in furnace	6	—
Glass melted	*c.* 20	—
Clay mixture	5	—

Method of Analysis

With so many variables involved, the only practicable way of distinguishing between the more essential and the less essential was to make out separate hand-punched cards for each pot and to record thereon the appropriate level of each factor. Hand sorting with the aid of a long needle made it possible to cross-classify at will and summarise different categories quickly.

Preliminary inspection of the data revealed that, in common with much other 'life' data, the distribution of pot lives is skew. In addition to the general skewness there was an unduly large accumulation of pots with lives of one day or less. Since these could be accounted for as having failed either before filling with glass or on their first filling, it seemed reasonable to regard them as being different from the normal run of failures and they were analysed separately. For the remainder whose lives varied between 2 and 100 days we had to decide whether or not to make a logarithmic or other transformation of the data. The details of the analysis had later to be explained to the technical staff involved and to the recording clerks who would use the method, and it is sometimes difficult to convince the statistical layman of the need or justification for such juggling. For this reason we decided against any transformation but used throughout the median of the pot lives in order to discount the effects of what we felt to be artificially long lives; in fact it will be seen that our method of ratios gives practically the same result as a logarithmic transformation.

In order to obtain a preliminary estimate of the relative lives under given conditions it was necessary to select from the total of 400 pots those portions of the data within which balanced comparisons could be made. For example Table II demonstrates how the initial comparison between makers was made. Thus, among covered pots used with town's gas in furnace A there were 9 pots made by Smith which had a median life of 35 days, 3 pots by Jones with a median life of 20 days and 2 pots by Robinson with a median life of 37 days, giving approximate ratios of Jones/Smith = 0·57 and Robinson/Smith = 1·06. Open

PLATE I. Glass-melting pots being made by building up a 'sausage' of clay into a spiral.

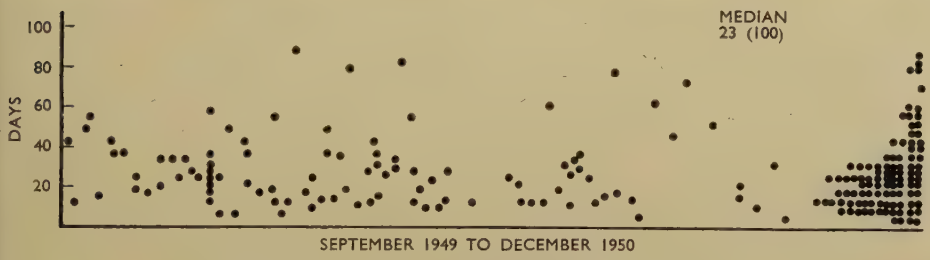


PLATE II. Life of individual pots made by Jones.

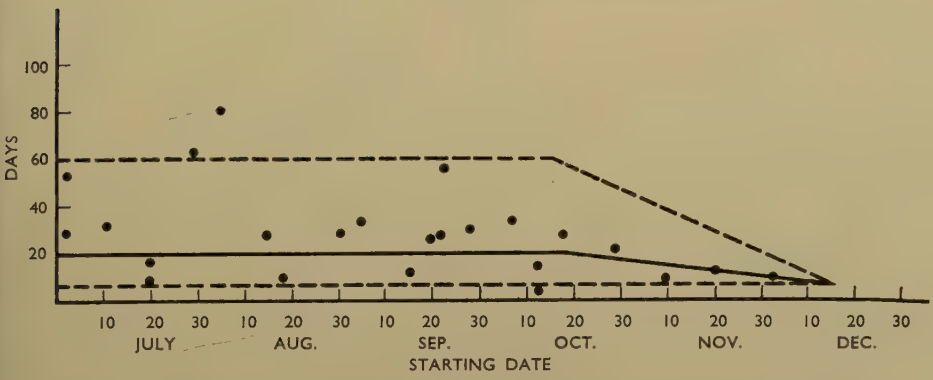


PLATE III. Lives of individual pots made by Smith. Sloping limits are necessary, as at a given date only short-lived pots are broken towards the end of the period and the others have not completed their lives. Nearly all pots fail within two months.

pots in the same furnace included 25 Smith pots with a median life of 15 days, 7 Jones pots with a median life of 15 days and 19 Robinson

TABLE II

Median lives in days of pots used under comparable conditions

Pot Type	Furnace Gas	Furnace	Pot Maker*			Ratio	
			Smith	Jones	Robinson	$\frac{J}{S}$	$\frac{R}{S}$
Covered	Town's	A	35 (9)	20 (3)	37 (2)	0·57	1·06
		C	50 (7)	50 (14)	57 (2)	1·00	1·14
	Producer	D	20 (4)	45 (2)		1·13	
Open	Town's	A	15 (25)	15 (7)	10 (19)	1·00	0·67
	Producer	B	25 (17)	20 (7)	15 (22)	0·80	0·62
		D	20 (4)	15 (3)	10 (1)	0·75	0·50
Half	Town's	C	12 (4)	5 (6)		0·42	
	Producer	B	10 (20)	10 (24)		1·00	
Weighted Ratio						0·82	0·80

* The figure in brackets indicates the number of pots on which each median is based.

pots with a median life of 10 days, giving the ratios J/S = 1·00 and R/S = 0·67. Other comparisons gave further estimates of the relative lives of Jones and Robinson pots. Since it was felt, in view of the large variability, that none of these estimates could be relied upon alone, it was reluctantly decided to abandon temporarily any thoughts of interactions in order to obtain more broadly based estimates of the main effects. Accordingly the ratios were averaged, using weightings appropriate to the number of pots involved in each. This gave weighted ratios of 0·8 for both J/S and R/S, from which it was concluded that, to a first approximation, Jones's and Robinson's pots lived, *ceteris paribus*, about four-fifths as long as Smith's.

By selecting other portions of the data it was possible to obtain similar weighted ratios for comparing furnaces, gases, and pot types. The resulting list of ratios is shown in Table III. By means of these ratios, assuming the effects to be multiplicative, it is possible to convert back to a standard life (the standard chosen being a Smith open-type pot used in C furnace with producer gas) the lives of pots made and used under different sets of conditions. For example, Table IV shows how to estimate the standardised lives of two pots with very different backgrounds. When the lives have been standardised to correct for the differing backgrounds it is found that pot X which, with the unfavourable combination of being a half-pot made by Robinson and used in B furnace, lasted only 3 days, was really no worse of its kind than pot Y which lasted for 37 days under very favourable conditions.

To avoid unnecessary repetition of arithmetic, further tables were drawn up (Table V) embodying all the corrections for a given set of conditions into a single factor by which the actual life is to be multiplied to give a standardised life, enabling comparisons to be made between individual pots with vastly different backgrounds. In addition, it is desirable to keep track of changes in the relation between any two makers, furnaces, etc. To do this it is necessary to correct for all other variables except the one concerned. The final column in Table V shows the combined factors, based on Table III, to be used for comparing pot types.

TABLE III

Pot History	Relative Life
Made by:	
Robinson	0·8
Jones	0·8
Smith	1·0
Used in Furnace:	
A	0·6
B	0·6
C	1·0
D	0·4
Type of Gas Firing:	
Town's	1·5
Producer	1·0
Type of Pot:	
Half	0·5
Covered	2·5
Open	1·0

TABLE IV

	Pot X	Pot Y
Maker	Robinson (0·8)	Smith (1·0)
Furnace	B (0·6)	C (1·0)
Gas	Producer (1·0)	Town's (1·5)
Type	Half (0·5)	Covered (2·5)
Actual life (days) ..	3	37
Standardised life (days)	$\frac{3}{0{\cdot}8 \times 0{\cdot}6 \times 1{\cdot}0 \times 0{\cdot}5} = 12$	$\frac{37}{1{\cdot}0 \times 1{\cdot}0 \times 1{\cdot}5 \times 2{\cdot}5} = 10$

Having standardised the lives in this way for all the data available, the next step was to obtain more precise estimates of the ratios by a process of iteration. For each maker, each furnace, each gas, and each pot type, the standardised lives were plotted in a time series according to the date of starting and, if no marked drifts or changes in level were noticeable, the points for each were collected up to form a frequency diagram as in Plate II. The median of this diagram was then taken as an improved estimate of the expected life of the pots concerned.

Thus in Plate II the median life of 100 pots made by Jones is 23 days; and similarly for every other variable. These improved estimates, in turn, enabled improved ratios to be drawn up, giving other tables similar to Tables III and V. By repeating the process three or four times until a further repetition gave an answer no different from the previous one, a final table of ratios was drawn up.

The first step, in which approximate ratios were obtained by comparing the few pots with common backgrounds, could be omitted, but, if it were, many more repetitions of the second step would be necessary.

TABLE V

Maker	Furnace	Gas	Factors for Comparing			
			Individual Pots			Pot Types
			Pot Type			
			Half	Open	Covered	
Jones Robinson	D	Producer Town's	6·2 4·2	3·1 2·1	1·2 0·85	3·1 2·1
	A, B	Producer Town's	4·2 2·8	2·1 1·4	0·85 0·55	2·1 1·4
	C	Producer Town's	2·5 1·7	1·3 0·85	0·5 0·35	1·3 0·85
Smith	D	Producer Town's	5·0 3·5	2·5 1·7	1·0 0·65	2·5 1·7
	A, B	Producer Town's	3·5 2·2	1·7 1·1	0·65 0·45	1·7 1·1
	C	Producer Town's	2·0 1·3	1·0 0·65	0·4 0·25	1·0 0·65

To compare individual pots, multiply actual lives by the appropriate factors in columns 4, 5, and 6. To compare pot types, multiply actual lives by the appropriate factors in the last column.

The finally corrected life for each pot was plotted on a control chart similar to Plate III with control limits of about 19/20 probability guessed from the frequency diagram. By reference to this control chart it can then be seen whether the general level of performance is remaining substantially the same or changing.

A similar chart is made for each maker, each furnace, each gas, and each pot type, by correcting for all the factors except the one being plotted. Plate III shows the control chart for Smith-made pots, with median at 20 days and limits at 4 and 60 days. Since each pot is plotted at a point on the time scale corresponding to its date of starting it is apparent that, at any given date, only those pots will appear on the chart which have failed up to that date. For this reason it is not to be expected that the more recent failures will straddle the median.

To facilitate the fairer assessment of these later points the device of sloping limits has been adopted. Since the lag involved is not likely to exceed 60 days, the median line is drawn to slope down to the lower limit over a period covering the most recent 60 days and the upper limit brought down to converge with these two lines at the same date.

It is not suggested that the above method is a quick one, but, once done, it has replaced the old method which involved computing the arithmetic mean of all pots in a furnace, including early failures, and has enabled much more valid and useful conclusions to be drawn from the same data.

The authors wish to thank Lemington Glass Works Ltd., Lemington-on-Tyne, for permission to include the data presented in this paper.

THE EXPERIENCE OF A USER OF A CONSUMER PANEL

ROBERT N. WADSWORTH
Cadbury Bros. Ltd., Bournville

Mr. Wadsworth is a firm believer in the value of consumer panels in market research, although he is aware of the difficulties associated with their use. This article is based on experience of the Attwood Consumer Panel, and deals with questions of accuracy and with the practical application of the results.

A CONSUMER panel can be said to be established when the same group of individuals is submitted to a series of intermittent interviews or is required to make a series of reports over a period of time. Thus the panel operation should be distinguished from the *ad hoc* survey or a series of surveys such as the Gallup Poll, when each of the group of people selected on each occasion is submitted to a single interview.

Basically, the advantage of a panel over the *ad hoc* survey in market research is that it makes possible a study of consumers' actions through time. It is a matter of common experience that memories are short and unreliable. When we have asked people whether they buy $\frac{1}{4}$ lb. of cocoa every week, once a fortnight, once a month, or less often, the answer is likely to correspond to three or four times the known rate of consumption. We cannot rely on housewives to tell us whether, although they use Bournville cocoa regularly, they did try a tin of Fry's two months ago. In fact there is a strong *prima facie* case for experimenting with the consumer panel in order to establish data on frequency of consumption and loyalty to a brand, which are not otherwise obtainable.

This case is strengthened if the panel is capable of producing trends and other market information usually obtainable by other means, for in this way further checks are obtained on established techniques, and ultimately it may be possible to treat the panel as a basic continuing service which only needs supplementing occasionally with special enquiries.

With this in mind my firm became a client subscriber to the Consumer Panel Service initiated in 1948 by Messrs. Attwood Statistics Ltd., to whom I am indebted for help in writing this paper. The object of this service is to give manufacturers of certain branded and nationally distributed consumer goods a statistical record of the purchases made by a representative panel of 2000 households, each reporting weekly on the diary form provided. Let me describe to you this panel service.

Setting up and Maintaining the Panel

Briefly, the method of obtaining a representative sample of households was as follows. First the population was divided into seven geographical areas and further subdivided into sub-areas consisting of urban districts, rural districts, and cities of various sizes. The selection of 94 towns and rural districts was then made deliberately to fit this stratification. In the London area a rather special procedure had to be followed: the area was divided into quadrants with inner and outer rings to give sections of nearly equal populations. The selection of the wards within each district was random, and when these were fixed the names of the households were drawn from the appropriate rating lists, taking only private dwellings. In rural areas, in order to bring the geographical spread within reasonable limits, all addresses outside a mile from a centre point in the district were discarded.

The considerations involved in this method of selection merit a paper to themselves, and this outline and the few comments I have to make must be brief. The general method seems to me to have been satisfactory except that it might have been better to select the wards as well as the districts on a purposive basis or, alternatively, to use a larger number of wards, since the class composition of a ward tends so often to be homogeneous.

On the basis of this selection nearly all the addresses were successfully visited, sometimes after many repeat calls. From about 90% of the addresses it was possible to secure data on family composition and to classify the household according to the age and social class of the housewife. It was then possible to compare the characteristics of the sample with the available demographic data as a general check on the efficacy of the method. Moreover, data were collected for reserve lists of households from which to provide the panel replacements.

I should now like to say something about the important question of co-operation and response. In the first place, it should be made clear that the household panel member is rewarded for her co-operation by a small payment every quarter, and that she is told of this at the original interview. As I have said, it was possible to secure family data from about 90% of the households visited. About 80% were enrolled on the panel, but there was an immediate drop in effective membership, as one in five of those enrolled failed to submit their first report. The proportion failing to complete the forms fell in succeeding weeks, but by the end of six weeks 60% of those originally enrolled remained in the panel—that is, 48% of those originally contacted. That, of course, creates a very serious flaw in the sample and immediately raises the question of replacement. The method of replacement is to select from the reserve lists a family with similar characteristics to the lapsed family so that the panel, as described by the demographic classifications established from the original panel selection, is not allowed to change.

The number required to be interviewed to ensure eventual co-

operation varied according to the age and income group of the housewife; generally speaking, the older and poorer groups were the more difficult to secure. The proportions falling out in the various groups, once enrolled, did not vary nearly so much. The better understanding of the upper income groups to some extent offset the lack of financial inducement; the financial attraction to the housewife with children to some extent offset the lack of time; and so forth.

It may well be argued whether the expense of this elaborate method of sampling was justified in view of the level of response achieved. I think it probably was justified, partly because it was possible to collect the control data, on the basis of which the future panel composition was to be maintained when making replacements; and partly because, in view of the prospective permanance of the panel, it was important that the initial selection should be done as well as possible by eliminating the serious possibility of investigator biases.

A good deal can be said on the question of the desirability of limiting the panel life of each household. It seems to me that the arguments depend on the uses to which the results are put. If one has no access to other market data there is a case for completely changing the panel from time to time in order to ensure that accidents in the original sample do not permanently colour the market picture. If, on the other hand, one can apply checks to the data and these are satisfactory, and if one looks mainly to the panel for trends, there is much to be said for continuity. This, however, is subject to the problem of conditioning. I was told the other day of a member of the panel in the pilot of the Government's household budget survey who had been so appalled when he saw the figures of what he was spending on cigarettes that he gave up smoking after three weeks' membership! Conditioning is a danger in all panel work, but as yet there is no evidence that it is a serious factor in the Attwood Panel. On balance, therefore, there seem to be strong arguments for maintaining continuity.

There is much also to be said on the questions of panel-member relations, on the form of diary, and on the fieldwork and editing procedures. Messrs. Attwood would, I know, acknowledge that they have benefited by drawing on American experience in the operation of household panels, e.g. the desirability of compiling the diaries in booklet form has been established. In addition, they themselves have developed by trial and error their own techniques of panel relations and their own design of diary report.

The field work is confined to the initial interview followed by a call a week later to elaborate any points which were not made clear at the first call. After these two calls contact is maintained as far as possible through the post.

On the question of editing the diaries the correct reporting of prices paid presents the biggest difficulty. Under existing shopping conditions, and as compared with pre-war, only a small proportion of housewives are able to report exactly the prices which they pay for

their purchases. In order that the purchases shall be correctly described before any summaries are tabulated it is necessary to correct prices reported and also to code the descriptions given. It is at this stage that any queries are taken up with the housewife. Usually these arise from an inadequate description of the purchases or from a failure on the housewife's part to give details of the retail outlet at which the purchase was made. About 100 postcards are sent out to housewives each week asking for further details. As more than 12,000 purchases are reported each week it will be seen that less than 1% of these have to be referred back to the housewife.

Before coming to the discussion of the panel results I should explain that these are presented in four-weekly reports, showing the quantities and values of the various brands purchased according to certain regional data. These reports are available only three weeks after the end of the four-weekly reporting period—an interval which is remarkably small in view of the processes which have to be performed. These monthly reports are supplemented by a long-period report—for our purposes we prefer a six-monthly report, whereas other firms may prefer a three-monthly one. This is not a summary of the four-weekly reports but contains analyses of purchases by all the demographic classifications used, and special analyses such as frequency of purchase, consumption a head, and 'brand loyalty,' for which the panel is peculiarly appropriate, since it allows of examination of the purchases made over a period of time, e.g. six months.

Tests of Accuracy

With all surveys the results should be approached critically, and it is important to devise as many cross-checks as possible. The first test we applied was to estimate from the panel results national figures for our own brands of beverages and to compare them with sales figures. We were fortunately able to do this when we were still in a sellers' market and when we had other evidence that there was little variation in shop stocks. It appeared that the panel substantially overstated the true figure. However, as experience was gained the comparison of the panel results adjusted for this tendency to overstatement gave a very fair indication of the market position as we proceeded into a buyers' market. More recent experience has justified our reliance on the panel figures as a basis of our sales estimates rather than our own sales figures, which have been affected by uneven buying by traders.

One of our first concerns after discovering this overstatement was to know whether we could rely upon the overstatement factor to be the same for the different brands in the various markets. Here we had a certain amount of check data. We had actual factory sales for more than one brand in one market, and also we had the results of other consumer surveys. Table I compares the relationship of the actual sales of two brands with proportions reported by the panel and suggests that the overstatement factor is approximately the same for both brands.

TABLE I

Comparison, for two brands, of factory sales and consumer purchases as estimated from panel results

Brand	Factory Sales (%)	Panel Results (%)
A	80·4	81·6
B	19·6	18·4
	100·0	100·0

A similar test was carried out comparing the ratio of sales of different packets with the purchases estimated from panel reports and a fair measure of agreement was obtained, as is shown in Table II.

TABLE II

Comparison of quantities of package sizes given by factory sales and consumer panel results

Size	Factory Sales (%)	Panel Results (%)
1	66·1	68·5
2	26·6	27·5
Other	7·3	4·0
	100·0	100·0

The comparison of sales a head in the seven regions with the panel results was not nearly so satisfactory. This arises in part from a fairly wide variability in consumption between districts within regions, and from the smallness of the regional samples. There is some degree of correlation, but the regional data are too unreliable to be of value for our purposes. The obvious solution is a much larger total sample if this were financially justifiable.

The foregoing comparisons more or less exhausted the most certain checks, e.g. those with our sales figures. We were, however, able to compare the pattern of purchasing households with the results of our *ad hoc* surveys. Although we were not able to establish any exactly comparable time periods between the panel surveys and our *ad hoc* surveys, the panel results seemed to confirm the pattern we had previously established.

Finally, as our one remaining check we arranged a postal questionnaire to panel members asking them to report what they had had to drink at various time intervals on the preceding day. These results were compared with those obtained in an *ad hoc* survey which we had separately arranged, and which had been very carefully designed to avoid the bias which arises from the informants being aware of the product in which the investigator is interested. There proved to be a close relationship between the two sets of results, as is shown by figures produced incidentally in the enquiry and given in Table III.

TABLE III

Percentage of adults drinking at various times

Time	Tea		Coffee	
	Ad hoc Survey (%)	Attwood Panel (%)	*Ad hoc* Survey (%)	Attwood Panel (%)
Before 8 a.m. ..	58	65	2	3
8–10 a.m.	60	53	5	6
10–12 noon ..	27	19	11	18
12–2 p.m.	53	48	11	15
2–4 p.m.	32	26	2	3
4–6 p.m.	73	76	2	1
6–8 p.m.	25	20	3	2
8–10 p.m.	20	20	9	11
10 p.m. and later ..	17	10	6	5

Percentage of children drinking at various times

Time	Tea		Milk	
	Ad hoc Survey (%)	Attwood Panel (%)	*Ad hoc* Survey (%)	Attwood Panel (%)
Before 8 a.m. ..	22	28	5	10
8–10 a.m.	50	36	20	25
10–12 noon ..	2	3	56	38
12–2 p.m.	37	22	12	14
2–4 p.m.	3	6	4	8
4–6 p.m.	61	61	13	19
6–8 p.m.	10	5	18	19
8–10 p.m.	7	4	8	7
10 p.m. and later ..	1	1	1	2

We did establish, however, that for adults the panel rates of consumption of our products were evenly some 25% or so above those recorded by our independent cross-section. There was some considerable divergence in the results recorded for children, but as we had to obtain this information indirectly in our *ad hoc* survey the results may not be so reliable. We believe the results from our 'once only' adult cross-section to be reasonably accurate, and this comparison seems to establish that at least a part of the panel overstatement is due to an inherent sample bias.

The Reasons for the Panel Overstatement

In spite of these tests we were still perplexed as to the principal reasons for the overstatement by the panel. It seemed to us that this could arise from a number of causes.

First, although the panel composition was consistent with the population stratifications according to the data available for the latter, it was possible that, in the process of replacing the large proportion of uncooperative housewives, a bias was introduced which was not measurable

by the classification system. It is not inconceivable that the type of housewife who drinks cocoa is the good steady type who would not mind keeping a household diary for a quarterly payment !

Secondly, it was possible that we were continually getting the benefit of new recruits who replaced the lapsed members and that these, in the early stages, tended to over-report. This, however, was 'not proven,' as the organisation stipulated the completion of at least one month's reports before qualifying for effective membership. It was thought that we might test the possible effects of non-co-operation by comparing results obtained from the households reporting regularly with reports of households who fell out within the first few weeks of signing up. The results of the test were unfortunately inconclusive, partly because the buying pattern of the product chosen was related to the phase of the ration period, partly because the sample of the households falling out was on the small side, and partly for the reason just mentioned that there was a known tendency to over-report in the initial report. I still believe, however, that a test on these lines could be carried out if the right products were selected and if the opportunity ever arose of a large intake of new panel members.

The possibility of a freak result was examined by taking the panel purchases over a six-month period and arranging the households according to quantities purchased. The main consumption of our products covered in this survey is concentrated in relatively few households, and it seemed possible that the freak inclusion of even a small excess of heavy consumers would appreciably distort the total result. We also arranged to identify the heavy consuming families, and some of these were specially visited by investigators to establish the probability of the reported rates of consumption. We concluded that the overstatement was not due to freak results.

It was possible that there would be a natural tendency to over-report an infrequently bought product, i.e. last week's purchase might inadvertently be again included in the current week. This was examined first by scrutinising the series of purchases and then by arranging to call back on a limited number of the panel members. These were cross-questioned on the rate of usage of the product, and in every case except one the truthfulness of the report was apparently established. Such a test, however, cannot be conclusively executed, and this factor may still be a contributory source of error. Nevertheless we are told that other subscribers to the panel, even one with a product which is bought with much the same frequency as ours, are satisfied with the accuracy of the results.

The Applications of the Panel Results

Consumer Purchases and Factory Sales. One application of our consumer panel results has been the comparison of our factory sales with consumer purchases. This, of course, can be done by means of a retail index, but for our products the panel results, once adjusted by the

exaggeration factor, give a good indication of the situation. This comparison at the present day is of special importance when wars and rumours of wars, and consequent shortages and fluctuations in prices, cause wide variations in consumers' and traders' expectations. It is most important for us to know whether an increase in sales is due to trade buying or to consumer buying. To complete the pattern we should of course know something about pantry stocks, but our panel operation has not yet extended to that field!

Brand Trends. It involves no stretch of the imagination to see how the same monthly aggregate information can be plotted for each brand to establish the trend in consumption. Again, this result is not exclusive to the panel operation; the monthly report is in many respects equivalent to a series of *ad hoc* studies of a population cross-section, although variations are not subject to the separate sampling errors that arise in the selection of successive cross-sections.

Studies of Rates of Purchase. When, however, we come to consider the long-term report—we decided, as I have said, to take half-yearly reports—we have a body of information which can be reliably obtained only by means of a panel. On the question of brand loyalty, for instance, we might find that over the period of the report only a small proportion of the housewives in one market buy more than one brand. In another field we might discover that the brand loyalty is not so strong and that people seem to like 'ringing the changes.' From this we would infer that the advertising approach to the two markets should be fundamentally different.

Again, we were able to study the frequency of purchases and to demonstrate to our advertising and marketing people among what a small proportion of households the bulk of the consumption of one of our products was concentrated. What we should like to do is to study in much more detail the backgrounds of these heavy consumers. Unfortunately, if we subjected them to special interviews we should probably condition them; and if we rely on the normal classification data and on the information of their press readership, which is periodically obtained through the panel, the sample is too small to draw inferences. Possibly, if ever the panel should be renewed, we should endeavour to interrogate some of these heavy users either by post or by personal interview.

Rates of Consumption. We have found in our experience with the panel that preconceived ideas that households of a certain class were heavier users had to be reversed, because, although this class bought more often, the total volume bought was in fact exceeded in another group.

Problems in Panel Research

I can now summarise the position as I see it. We have sufficient experience to say that the consumer panel is a promising instrument of research and that, for some products at least, it gives much more

reliable information on rates of purchasing than a series of *ad hoc* surveys depending upon the individuals correctly recalling their purchases. It is also a much better means for studying the change in buying and spending habits and is capable of yielding data not otherwise obtainable.

Inherent in the technique, however, are certain difficulties. First, there is the question of choosing a representative sample of the population and, more difficult still, of obtaining co-operation and ensuring that the co-operating members of the panel remain representative. A part of this problem remains a study in human relations. The question arises of the elimination of sample bias. This may be unrevealed by application of the usual quota sampling tests in which the sample is checked by reference to regional information concerning age and income groups, and family composition. It seems to me that a possible solution is to be found in the selection of other criteria for the sample cells, so that there may be more methods of testing the sample biases introduced by non-co-operation and panel mortality. Co-operative Society membership is one type of classification; religious practices might be another if we had the relevant census data !

The problem of co-operation and effective response is much more important where some positive action is required of the respondents, such as the keeping of a diary, than when it is merely a matter of arranging for the same people to be subjected to a series of intermittent interviews. One is confronted with the same problem in dealing with panel mortality as in the initial stages of securing effective response.

Then there is the problem of the conditioning of the panel members and the introduction of a cumulative bias. This may arise from the development of an unnatural consciousness which may affect buying habits and choice of brand. It is a danger to guard against but is probably much less important in a panel based on a weekly diary of a selected number of household purchases than in one depending upon a series of personal interviews.

As with all consumer surveys, there is the difficulty of designing the questionnaire. A diary presents peculiar problems because, in order to indicate a field, the definition may sell the product. There is danger in having too few products in the diary, and there is danger in inducing fatigue and ineffective response if too many are used.

Finally, there may be problems inherent in the subject matter of the enquiry itself, i.e. in the products. Accuracy of response is likely to vary according to the nature of the product, the frequency with which it is bought, its cost in relation to the household budget, and the persons in the household most likely to use it.

I trust that, because I have dwelt on the practical problems of the panel technique, I have not induced a general scepticism on the efficacy of this approach to research. My hope is, of course, by the sharing of experience to arrive at a solution to some of the difficulties,

and any suggestions that may be made will be most welcome. The fact that we continue to subscribe to this service is perhaps the best evidence that we are not overawed by the difficulties encountered, and that we are already getting valuable market data. I am sure that ultimately there is an important future in the use of consumer panels, not only of the type that I have described but also of the type depending on actual interviews.

THE ACCURACY OF SAMPLING COAL

JACK HEBDEN *National Coal Board (NE Division)* and
GEOFFREY H. JOWETT *University of Sheffield*

In sampling a week's output of 300 wagons of coal in order to estimate the mean ash content, how many visits should be paid to the colliery each week? How many wagons should be sampled on each visit? How many 'increments' should be taken from each wagon? What errors are introduced by the techniques of 'reduction' and analysis? This article describes some experiments made to provide information on questions such as these, and discusses the application of the results.

THE typical colliery of the National Coal Board produces simultaneously a number of grades of coal, each of which is usually loaded into railway wagons by some form of conveyor. The quality of any particular grade varies with time; hence for quality control and for information about the nature of current production it is the practice to sample at fairly regular intervals. The frequency of sampling and the technique vary from area to area and from grade to grade. In the North Eastern Division the sampling of commercial grades of coal is carried out by teams of samplers, each team being based on an area laboratory and using a 30 cwt. van for transport. Samples have to be collected over a wide area and the pattern of visits to the collieries varies according to circumstances. A typical day's sampling would start at 8 a.m.; the van would set out with four men, would drop two at colliery A and then proceed to colliery B. On completion of the sampling at colliery B the van would return to colliery A to pick up the other two men with their samples, and would then go on to other collieries. Finally the van would return to the area laboratory with the samples. The number of visits in the day is limited, partly by the distances apart of collieries and laboratory, and partly by the number of grades to be sampled. Transport within a colliery is not usually easy, and heavy samples may have to be carried from sampling point to van by long and tortuous routes, as a glance at Plate I will show.

For uncleaned slack with between three and fifteen per cent. of ash, the grade considered here, a wagon load is taken as the unit of production. This unit is sampled by taking shovelfuls or 'increments' at regular intervals from the stream of coal as it flows to the wagon by conveyor belt. The increments are so arranged that the full cross-section of the stream is represented. The size of increment is about $2\frac{1}{4}$ lb., and all the increments—sixteen for two of our experiments—are collected together in a sack to make a bulk sample weighing about 36 lb., which purports to be representative of the contents of the wagon.

The next process is one of reduction (see Plate II), whereby a small and representative quantity of finely ground coal is obtained from the bulk sample for analysis in the laboratory. The bulk sample, crushed to give 90 per cent. passing a one-eighth inch square-mesh sieve, is passed through a static circular riffle which selects a quarter of it; this is halved by means of a standard symmetrical riffle and, after further crushing and mixing, about three ounces are removed to a bottle from the circular tin holding the accepted half; from this the analyst removes a one-gram sample by spatula and analyses it. Among other variables he determines the ash content, to which attention will be confined in this paper. During the whole process of sampling many precautions are taken to secure good mixing and freedom from bias.

Since a colliery is visited only two or three times a week, the resulting ash determinations are very limited in the extent to which they can represent the week's output. They are used, not so much to detect abrupt changes of quality, a task for which much more intensive sampling would be needed, as to give an indication of the quality level and to detect undesirable trends in it. This indication need not be very accurate, though it should be free from bias.

The following experiments were designed to measure the accuracy of the three stages of sampling: (*a*) sampling wagons from a stream of wagons, (*b*) sampling increments from a wagon-length of conveyor belt, and (*c*) reduction and analysis. They were to provide a factual basis for criticism of present sampling practices and for the rational discussion of alternatives. If, for example, the colliery is visited l times a week and on each occasion a systematic sample of m increments is taken from each of n consecutive wagon-lengths of belt, what is the best choice of l, m, and n? If some energy is to be devoted to improving accuracy, to which point is attention most profitably directed? If economy dictates a reduction in sampling, where are the cuts best made? The answers to these questions depend on the relative sizes of error introduced at each of the three stages of sampling, as will be shown below. These stages were investigated separately in three experiments which were carried out incidentally, and therefore at little cost, during a heavy sampling programme designed for another purpose.

Wagon Stream Sampling Experiment

Because of the limited availability of samplers, the design of this experiment had to be very flexible. Wagons were sampled (using sixteen-increment samples) in consecutive pairs, for a pair could be done in a single visit; visits were deliberately made at irregular intervals. Thus the very important comparison between consecutive wagons was a good one, and comparisons were also available between wagons with widely varying degrees of separation. Results were obtained for eighty-two pairs of wagons.

Analysis of the resulting data involves a measure termed the *mean semi-squared difference* (MSSD), which is denoted by $\delta(\tau)$ and is defined

PLATE I. Typical colliery layout, showing difficulties of transport of bulk samples. *A* is the entrance to the sampling point, *B* are two streams of wagons containing different grades, and *C* is a 30 cwt. van parked as near to the sampling point as possible.

PLATE II. Apparatus for the reduction of bulk samples. *A* is the crusher, *B* the circular riffle, *C* the symmetrical riffle, *D* the small crusher, *E* a circular tin, *F* a tube, and *G* a bottle.

as the average value of one-half the squared difference in ash content of wagons separated by an interval τ. Wagons which are close together in the stream tend to differ less than widely separated wagons; in consequence $\delta(\tau)$ usually increases with τ. For a random series, where there is no such tendency, $\delta(\tau)$ $(\tau \neq 0)$ is constant and equal to the variance of the series, a fact which is partly the reason for halving the squared differences to obtain $\delta(\tau)$. The curve of values of $\delta(\tau)$ plotted against τ is called the MSSD curve and may be estimated by graduating actual average semi-squared differences calculated from an observed series. Using a simple graphical technique described elsewhere[1] MSSD estimates were obtained from the eighty-two pairs of ash contents and are illustrated in Fig. 1; the points shown are actually average MSSD's, taken over convenient intervals of τ and corrected for the error of sampling a wagon. The MSSD curve was filled by eye; it measures fundamental properties of the variation and shows that while wagons near together differ in ash content, those far apart differ more.

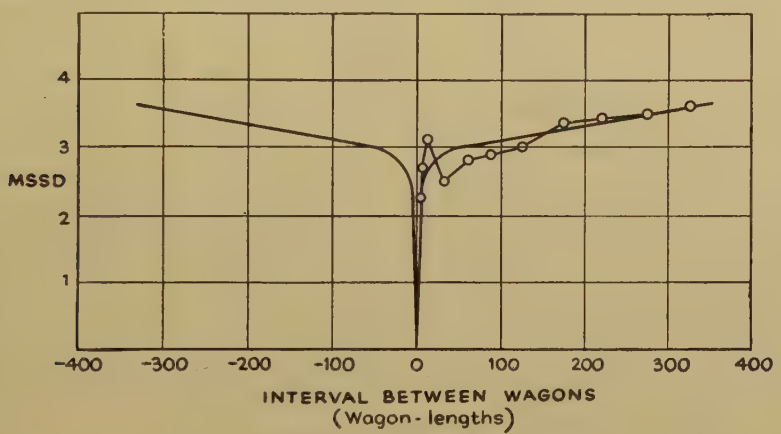


FIG. 1. MSSD curve for the wagon stream.

The week's production of the grade is approximately three hundred wagons. Table I shows the accuracy with which the true mean ash content of these is estimated by the mean ash content of various kinds of systematic sample. This was determined from the MSSD curve by a technique described by one of us (GHJ) in an earlier paper,[2] and which is illustrated in Appendix (i). It will be seen from the table, for example, that if the mean weekly ash content is estimated from samples from the two widely separated wagons 75 and 225, the error variance is 1·35 and less than the value when the wagons are consecutive and numbered 150 and 151.

Conveyor Belt Sampling Experiment

In this experiment samples of seven increments were taken in a special manner from each of twenty-four wagon-lengths of belt. In Fig. 2 the section of belt corresponding to the contents of a wagon is

illustrated schematically. For wagons 1, 4, 7, . . ., 22, increments were taken as shown in Fig. 2; they were evenly spaced along the belt except for the fifth, which was taken half an interval early; the second and third were taken from the left and right of the belt respectively and the remainder were taken from the centre. The increments were analysed separately, duplicate laboratory samples A and B being taken

TABLE I

Error variances of means of systematic samples from a week's production of 300 *wagons, excluding errors of reduction and sampling from the belt*

Serial Numbers of Wagons in Sample	Error Variance of Sample Mean
..... 150	2·80
..... 150, 151	1·67
.....75.....225.....	1·35
.....50..... 150 .....250.....	0·89
.....75, 76.....225, 226.....	0·79
.....50, 51.....150, 151.....250, 251.....	0·52
.... 25.... 75 .... 125 .... 175 .... 225 ....275	0·44

from each of the increments 2 and 3 to furnish an estimate of variance of the error (peculiar to the experiment) due to reduction and chemical analysis. For the other wagons the arrangement used was similar, except that left, right, and centre positions were interchanged; thus for wagons 2, 5, 8, . . ., 23, increments 2 and 3 were taken from the right and centre positions, the other five increments being taken from the

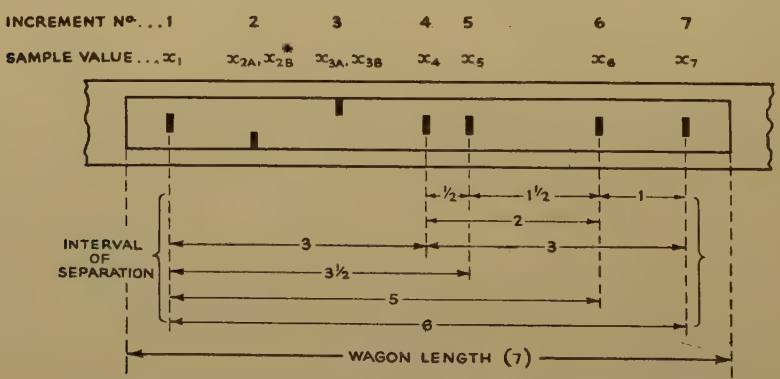


FIG. 2. Arrangement of increments in the conveyor belt sampling experiment. [* A and B readings are duplicates taken from the same increment.]

left; and for wagons 3, 6, 9, . . ., 24, increments 2 and 3 were taken from the centre and left positions, the other five increments being taken from the right. The arrangement of increments 1, 4, 5, 6, and 7 along

the belt was designed to give comparisons between increments separated by as many different intervals of τ as possible (namely $\frac{1}{2}$, 1, $1\frac{1}{2}$, 2, $2\frac{1}{2}$, 3, $3\frac{1}{2}$, 5, and 6 units of one-seventh of a wagon-length) so as to give MSSD estimates for many values of τ. These estimates were to be obtained by averaging semi-squared differences over the twenty-four wagons; thus, for example, the average value of $\frac{1}{2}(x_5 - x_7)^2$, corrected for the inflating effect of reduction error, would furnish an estimate of δ $(2\frac{1}{2})$. Increments 2 and 3 were to be compared with the mean of increments 1 and 4 to give information on the nature of variation across the belt; and the three permutations of arrangement were designed to ensure that the MSSD curve would be descriptive of the average behaviour of ash content of coal on the left, centre, and right sections of belt instead of merely the centre section. Although the experiment was complicated from the statistical point of view, it was not unduly troublesome to carry out since the uniform spacing of increments was disturbed in only one case.

The results of the analysis of variation across the belt are given in Table II; from these it was concluded that ash content in the centre of the belt is consistently greater than at the sides. Moreover, apart from such a constant bias, there was an apparent variation from wagon to wagon in the centre/sides bias giving rise to the observed 'variance of bias' given in Table II. It is shown in Appendix (ii), however, that this was apparent rather than real, and that the variation was of the same kind as that along the belt. The results of the analysis of variation along the belt are illustrated in Fig. 3, where the MSSD estimates and curve are shown.

TABLE II

Variations across the conveyor belt

Wagon Numbers	Mean Ash Contents			Mean Bias (centre minus mean of sides)	Observed Variance of Bias (variation of bias between wagons within the group)	Hypothetical Variance of Bias (from provisional MSSD curve, Appendix (ii))
	Left	Centre	Right			
1, 4, ..., 22	5·419	6·919	6·119	1·150	5·21	2
2, 5, ..., 23	7·069	7·069	6·138	0·466	2·41	2·7
3, 6, ..., 24	5·550	6·606	5·150	1·256	0·91	2·7
Mean ..	6·013	6·865	5·802	0·952 (s.e. = 0·344)	2·84	(2·5)

In routine sampling, increments would be taken from a wagon-length of belt at regular intervals and from left, centre, and right sections in turn, as illustrated in Fig. 4. They would then be bulked, reduced, and analysed, following the procedure described below in the reduction experiment. The error variance of the mean of such a

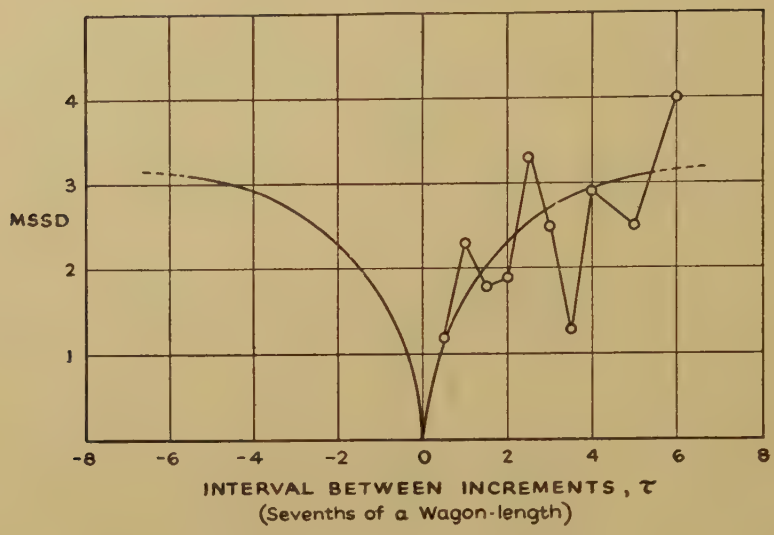


FIG. 3. MSSD curve for the conveyor belt.

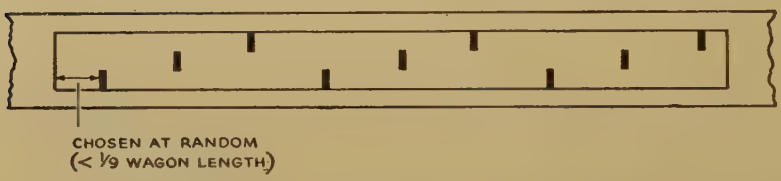


FIG. 4. Typical arrangement of increments in routine sampling of the conveyor belt.

TABLE III

Error variance of means of systematic samples from the conveyor belt (excluding reduction error)

Number of Increments per Wagon-length (n)	Error Variance of Mean
2	0·761
3	0·434
4	0·256
5	0·193
6	0·131
9	0·068
12	0·036
15	0·025
18	0·015
21	0·007

NOTE. It is assumed that the first of the n increments is taken at random in the first $1/n$th part of the wagon-length of belt.

sample may be deduced from the MSSD curve and is given in Table III for different numbers of increments. Details of the statistical technique are given in Appendix (ii).

Reduction Experiment

In this experiment standard laboratory samples and controls were taken from bulk samples from thirty-five wagons. The process of reduction of each bulk sample is illustrated schematically in Fig. 5;

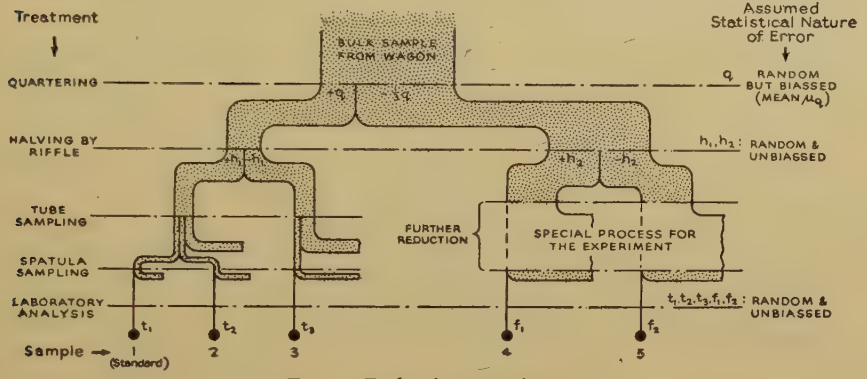


FIG. 5. Reduction experiment.

sample 1 is the standard sample obtained by the normal routine method of reduction, and samples 2–6 are controls taken especially for the experiment. The form of reduction or treatment at each stage of the process is indicated on the left of the figure; symbols representing the errors introduced at each stage are shown at the divisions of the 'stream'; and a description of the assumed statistical nature of these errors is given on the right. The errors vary from one bulk sample to another, having variances and in one case (q) a non-zero mean.

In designing the experiment it was assumed that h_1 and h_2 had the same variational properties, and the errors of tube sampling, spatula sampling, and laboratory analysis were considered not separately but as a single complex error. There were thus four types of error entering into the experiment, characterised by the symbols q, h, t, and f. Four independent comparisons of ash content could be made between the five samples from each bulk sample, and the design of the experiment was such that the variances (and for q the mean) of the four types of error could be deduced from these comparisons. This experimental arrangement has an advantage over the more conventional approach of multi-stage sampling (e.g. take two replicate samples from each of the four main streams) in that the four types of error were investigated equally thoroughly, instead of having more comparisons at the later stages of reduction (e.g. one for quartering error, two for halving error, and four for tube-spatula-analysis error for every bulk sample); it is thus much more economical. Unfortunately it is not so amenable to analysis by the ordinary analysis of variance, and standard errors must

in general be used to assess the significance of estimates; yet for estimation, and for assessing the accuracy of estimates, it is no less convenient, particularly because of the asymmetry introduced by quartering.

Although the analyst's results were believed to be consistent from day to day, a check on this was carried out by analysing the bulk samples in groups, doing samples 1, 3, and 5 on one day and samples 2 and 4 on another. This check proved the analyst to have a bias which varied from day to day; it was, however, possible to eliminate this from the comparisons by a straightforward method, which in the interests of brevity will not be described here.

Estimates of the mean and variance of q and of the variance of h, t, and f are given in Table IV. In most cases the standard errors

TABLE IV

Estimates of variation parameters for the four types of error

Nature of Error	Estimated Mean	Estimated Variance
Quartering (q)	— 0·02 (s.e. 0·03)	0·002 (s.e. negligible)
Halving (h)	—	— 0·013 (s.e. 0·0065)
Tube sampling, etc. (t) ..	—	0·066 (s.e. 0·0158)
Further reduction (f) ..	—	0·078 (s.e. 0·0180)

of the estimates are also given; for q, however, the estimate of variance was negligibly small, and the calculation of its standard error, which would be comparable to the others in size, was not considered worth while; furthermore, the estimate of the mean of q failed to exceed its standard error. Further details of statistical technique are given in Appendix (iii).

Practical conclusions resulting from Table IV were that only the tube-spatula-analysis error t is important; the errors q and h could be negligibly small, and the remaining error f, while comparable in size to t, is of no importance outside the experiment. Thus excluding analyst's day-bias, which it is hoped to remove, the variance of the error of reduction and analysis is effectively 0·066: the variance of t.

Application of Experimental Results

The results of the three experiments may be combined to give the properties of various sampling plans, which may be examined and compared in relation to the practical organisation of sampling. It is rarely possible to apply any procedure rigidly but an idea of comparative merits assists in the choice of procedure ultimately laid down.

Suppose that two visits are made to the colliery during the week, say about a quarter and three-quarters of the way through, and on each visit three increments are taken from each of two successive wagons. The four resulting bulk samples are then reduced and analysed like the

standard sample in Fig. 5. If $\bar{x}$ is the actual mean of the four ash content determinations and μ the true mean ash content of the 300 wagons which form the week's production, $\bar{x} - \mu$ = error of sampling the stream of wagons (from Table II the variance of this is 0·79) + mean of four errors of sampling the belt by three increments (from Table V the variance of each is 0·434) + mean of four errors of reduction and analysis (from Table IV the variance of each is 0·066). Hence the variance $\sigma_{\bar{x}}$ of $\bar{x}$ about μ is given by

$$\sigma_{\bar{x}}^2 = 0{\cdot}79 + \tfrac{1}{4} \times 0{\cdot}434 + \tfrac{1}{4} \times 0{\cdot}066 = 0{\cdot}91$$

and $2\sigma_{\bar{x}} = 1{\cdot}91$. The observed mean $\bar{x}$ will thus seldom be in error by more than 1·91. Most of its error variance is due to sampling the stream of wagons; errors of reduction make only a trivial contribution. The variance of analyst's day-bias is not taken into account, but its contribution would in fact be comparable with that of the errors of reduction.

In Table V are given examples of a number of realistic sampling plans together with the error of the determination of mean ash content resulting from the use of each, computed in the manner illustrated above. Example 1 is typical of present practice. Example 2 is a programme of a type that has been used in the past for relatively intense sampling under exceptional circumstances.

TABLE V

Accuracy of possible sampling plans (examples for an uncleaned smalls produced at a rate of 300 wagons per week)

Example	Number of Visits per Week	Number of Consecutive Wagons sampled per Visit	Number of Increments per Sample	Error Variance of Mean of Week's Samples	Limits of Error
1	2	1	16	1·39	± 2·4
2	6	1	16	0·45	± 1·3
3	1	2	3	1·92	± 2·8
4	2	2	3	0·92	± 1·9
5	3	2	3	0·60	± 1·5

A change from the practice of example 1 to that of example 3 would result in only a slight decrease in accuracy but an enormous reduction in labour; only one visit a week would be necessary, and with bulk samples of only three increments (about 7 lb.) it might be possible for one man to cope with the sampling. On the other hand, a change to the procedure of example 4 would mean no increase in the number of visits; yet there would be the same reduction in the labour of handling bulk samples as before and a slight improvement in accuracy. Similar remarks apply to a change from the procedure of example 2 to that of example 5.

It is tempting to set about the task of constructing and minimising cost functions, but certain facts must not be forgotten. Firstly, the samplers have other coals to sample, which differ in their properties; for washed smalls, ash content may be expected to vary more smoothly along the belt, and while this would justify the taking of only a few increments per wagon, it would reduce the advantage of sampling two consecutive wagons per visit. Secondly, even if this were the only coal

TABLE VI

Comparisons between samples related to errors

	Difference	Structure of Difference in Terms of Errors	Mean Value of Difference	Mean Squared Difference
1	$\frac{1}{4}(x_1 + x_2 + 2x_3) - \frac{1}{2}(x_4 + x_5)$	$\frac{3}{4}q + \frac{1}{4}(t_1 + t_2 + 2t_3) - \frac{1}{2}(f_1 + f_2)$	$\frac{3}{4}\mu_q$	$\frac{9}{16}\{\mu_q^2 + \text{var}(q) + \frac{8}{9}\text{var}(t) + \frac{1}{2}\text{var}(f)\}$
2	$\frac{1}{2}(x_1 + x_2) - x_3$ ..	$2h_1 + \frac{1}{2}(t_1 + t_2 - 2t_3)$	0	$4\,\text{var}(h) + \frac{3}{4}\text{var}(t)$
3	$x_1 - x_2$	$t_1 - t_2$	0	$2\,\text{var}(t)$
4	$x_4 - x_5$	$2h_2 + f_1 - f_2$	0	$4\,\text{var}(h) + 2\,\text{var}(f)$

involved, construction of a cost function would be difficult; if the sampler can visit a particular two collieries in a morning and has time to spare at each, he might as well sample two consecutive wagons as one; but if he is pressed for time, the extra wagon may cut out the second visit. The collieries are by no means equidistant from the laboratory or from one another; nor is it much use visiting a colliery for sampling at a change-over of shifts. In spite of these snags the results definitely point to a revision of sampling procedure; some reduction in the number of increments per wagon seems desirable and it might be desirable, in suitable circumstances, to consider the possibility of sampling consecutive trucks. We are thus far on the road to determining 'the best values of l, m, and n.' It also seems that there is no point in elaborating the reduction procedure to reduce the unbiased errors of reduction; it might even be simplified provided that biased errors are not thereby introduced.

Thanks are due to Mr. J. F. Scott, who did most of the computing, and to the National Coal Board for permission to use the data relating to ash contents of coal. The views expressed in the paper are those of the authors, and do not necessarily represent those of the National Coal Board.

REFERENCES

[1] Jowett, G. H., and Scott, J. F. (1952). 'Some Graphical Methods of Calculating Serial Correlations and Mean Semi-squared Differences.' Unpublished paper.

[2] Jowett, G. H. (1952). 'The Accuracy of Systematic Sampling from Conveyor Belts.' *Applied Statistics*, **1**, 50.

APPENDIX: DETAILS OF STATISTICAL TECHNIQUE

(i) *Wagon Stream Experiment*

By graphical analysis,[1] an MSSD estimate was obtained for $\tau = 1$ and average MSSD estimates for the ranges $\tau = 2$–8, 9–17, 51–75, 76–100, 101–150, 151–200, 201–250, 251–300, and 301–350, the units being wagon-lengths of belt. These were corrected for conveyor-belt sampling and reduction errors and plotted against the mid-points of their respective ranges, as shown in Fig. 1; the MSSD curve was then fitted by eye.

To obtain the error variance of the mean ash content of, say, wagons 75 and 225 from a stream of 300 wagons, the procedure was as follows. The ordinates of the graduated MSSD curve between -300 and $+300$ were multiplied by the factor

$$\begin{aligned} 1 + |\tau|/300 &\text{ for } 0 < |\tau| < 75, \\ |\tau|/300 &\text{ for } 75 < |\tau| < 225, \text{ and} \\ -1 + |\tau|/300 &\text{ for } 225 < |\tau| < 300. \end{aligned}$$

Next, the area under the resulting curve was determined and divided by the interval (150) between the sampled wagons in question to give a value of 5·86. From this was subtracted the sum of the ordinates at $\tau = \pm 150$, the resulting figure being divided by the sample size (2) to give an error variance of 1·35.

To determine the approximate error variance of the mean ash content of two groups of r successive wagons centred roughly at 75 and 225, a correction of 2·25 $(\frac{1}{2} - 1/2r)$ was subtracted from the above error variance; for $r = 2$ this yields the value of 0·79 given in Table I. This correction was obtained by the rough procedure of treating the r successive ash contents as if they varied randomly with a variance of 2·25.

(ii) *Conveyor Belt Sampling Experiment*

Although the observed wagon-to-wagon variances of bias quoted in Table II may be expected to differ for the three groups of wagons because of the different arrangements of increments, the equal numbers in the three groups justify the calculation of the standard error of the overall mean bias in the usual way as $\sqrt{(2{\cdot}84/24)}$. The hypothetical variance of bias was deduced from the MSSD curve by treating the four increments as if they all came from the same section of belt; for example, for wagons 2, 5, and 8 this may be shown to be

$$\tfrac{3}{4}\sigma_0^2 + \tfrac{1}{8}(10\delta(1) + 2\delta(2) - \delta(3))$$

where σ_0^2 is the variance of errors of reduction and analysis estimated by the MSSD between duplicate samples.

If the hypothetical variances were the true values corresponding to the observed variances, $7(5{\cdot}21/2 + 2{\cdot}41/2{\cdot}7 + 0{\cdot}91/2{\cdot}7)$ would be a value from the χ^2 distribution with 21 degrees of freedom. Its value is

26·95, which is less than the 5 per cent. point. Thus the variation in the estimates of bias for the separate wagons was no more than could be accounted for by the spatial separation of the relevant increments and the nature of the variation along the belt.

If the bias had in fact proved to vary significantly from wagon to wagon, the provisional estimate of the MSSD curve would have been final. As such was not the case, the precision of the MSSD estimates was improved by using comparisons between all seven increments, with due allowance for the established bias of the centre, and for error of reduction and analysis. For example, we have as a supplementary estimate of $\delta(1)$, to be combined with the previous one, the expression

$$\tfrac{1}{2} \text{ var } (\bar{x}_1 - \bar{x}_2) - \tfrac{3}{4}\sigma_0^2$$

where the term in σ_0^2 is the appropriate correction for errors of reduction. If bias were not truly constant these supplementary estimates would tend to be a little too high, but the risk of this was regarded as amply compensated for by the reduction of sampling error resulting from the combination of estimates. The final estimates were obtained by averaging all estimates with the same interval as argument, and correspond to the points shown in Fig. 3; the final MSSD curve was fitted by eye.

The systematic sampling procedure illustrated in Fig. 4 is equivalent to combining three staggered linear systematic samples, one from each of the left, centre, and right sections of belt. The equal representation of these streams takes care of the constant part of the variation across the belt; moreover, since MSSD's between points in different streams may be taken as nearly the same (except for very small values of τ comparable with the width of the belt) as MSSD's between points in the same stream, the method described in the earlier paper[2] is approximately applicable and may be used to predict the accuracy with which the true mean ash content of the wagon is estimated by systematic samples with various numbers of increments. For example, to obtain the error variance of the mean ash for a systematic sample of nine increments the procedure was as follows. The ordinates of the graduated MSSD curve between -7 and $+7$ were multiplied by the factor $1 - |\tau|/7$. The area under the resulting curve was then calculated as 2·141; this was subtracted from the sum of the ordinates of the curve at $\tau = \pm 7/9, \pm 14/9, \pm 21/9, \pm 28/9, \pm 35/9, \pm 42/9$, and $\pm 49/9$, and the result was finally divided by the sample size (9) to give an error variance of 0·068.

(iii) *Reduction Experiment*

There is more than one valid way of deducing the four independent error parameters from comparisons between the five samples, but the way shown in Table VI is probably the best. The table shows how the differences in the left-hand column, in mean or mean square over the thirty-five bulk samples, may be used to provide estimates of measures

of the four types of error. Equating the mean squares to their theoretical expectations (shown in the right-hand column) and the mean of the first difference to its expectation provides five simultaneous equations in the five unknowns from which the estimates given in Table IV are readily obtained. The standard errors of these estimates, also given in Table IV, were calculated by large-sample methods on the assumption that all errors were normally distributed. For example, an estimate of the variance of the halving error h is given by

$$\tfrac{1}{4}\,\mathrm{Av}\,[\tfrac{1}{2}(x_1 + x_2) - x_3]^2 - \tfrac{3}{16}\,\mathrm{Av}\,(x_1 - x_2)^2,$$

where Av signifies the average of the term following. The two average squares in this expression are distributed independently as χ^2 with 35 degrees of freedom; hence the square of the standard error of the estimate is approximately

$$(\tfrac{1}{4})^2(\tfrac{2}{35})\{\mathrm{Av}\,[\tfrac{1}{2}(x_1 + x_2) - x_3]^2\}^2 + (\tfrac{3}{16})^2(\tfrac{2}{35})\{\mathrm{Av}\,(x_1 - x_2)^2\}^2$$

ESTIMATING INDIVIDUAL EXPENDITURE FROM FAMILY TOTALS

W. F. F. KEMSLEY
The Social Survey

Surveys commonly give estimates of the total expenditure of various types of families and it is sometimes desired to deduce from these estimates of the share to be apportioned to typical individuals. This is the problem dealt with in this article. Mr. Kemsley describes and discusses several methods, which have a wider application than to estimates of expenditure, and gives some results which are interesting in themselves.

Introduction

This is a subject of considerable importance because many inquiries provide only family totals. Examples can be found in inquiries into nutrition and the consumption of food. Sometimes this restriction is inevitable, as, for example, where consumption is derived from quantities purchased for the family as a whole, taking account of stock changes. Sometimes it is a matter of convenience in order to get sufficient cooperation from the selected group, since a request for separate information from each member of a family may lead to difficulties.

Family studies of this type can, however, be made to yield estimates of individual consumption. One method that has been used[1, 2] relies on comparisons of groups of families of identical composition and structure except for the presence or absence of one member. The procedure, described as the 'subtraction method' in the present paper, has the advantage of simplicity but it is inefficient. It depends on the segregation of groups of families of particular composition thus excluding many not conforming to one of the selected standard groups. The results are based on a part of the sample only and therefore are not as accurate as they would have been had a more efficient statistical procedure been used.

Better and more efficient is a method based on the principle of least squares. A procedure of this kind was employed by the author in 1945–6 while working in the Ministry of Food on the food budget surveys carried out by that department, and some results obtained at that time are given in the last section of this article. Subsequent reflection suggested that there were several alternative procedures, all based on the principle of least squares. These are discussed here and compared with results from the subtraction method using data obtained from certain surveys carried out for the Central Statistical Office by the Social Survey.

However, whatever method of estimation is used, some care is required in interpreting the results. The procedure implies the comparison

of families of different numbers and composition, but families of different size will tend to have different standards of living simply because they contain different numbers of persons. For example, a family which suddenly acquired an additional member would usually have its standard of living changed, although the precise effect would differ according to circumstances. The requirements of an additional child would usually have to be met out of the total resources of the family, which would remain unchanged apart from the payment of family allowances. An additional earner, on the other hand, may or may not improve the standard of living of the rest of the family. By the same argument, families of different sizes will, in general, also have different standards of living.

No adjustment has been made in this paper for the 'income effects' just described. The present estimates represent, therefore, average expenditure or consumption associated with the presence in the family of a particular type of person. For example, consider the case of children of a particular age group; there is evidence that if estimates are made separately for an only child and for children in families containing two or more children, the estimate for the only child is higher. Some of this difference at least will be caused by the lower standard of living of families with several children. The estimates obtained when no correction is made for the standard of living are therefore averages weighted implicitly by the distribution of standards of living actually found in the population represented by the sample.

The decision as to whether an adjustment should be made depends on the purpose for which the estimates are required. Moreover, adjustment is only possible if some family characteristic can be found which will give a reasonably accurate assessment of the standard of living. One method has been used by Mr. Nicholson,[3] although it was applied only to families of a simple structure. The estimates in his paper are based on a subtraction procedure and, although modified by the adjustment for the standard of living, they suffer from the deficiencies described here. Some combination of his regression technique with one of the least square procedures described below would be more satisfactory. Another method of assessing the standard of living which has been suggested[4] is dependent on a knowledge of the prices of certain items. In practice, neither method may be easy to apply, since it is not always possible to get the additional data.

Whether an adjustment for the standard of living is made or not, the resulting individual estimates do not necessarily give the actual share of the individual in the consumption and the expenditure. Without information relating directly to the individual it is not possible to be certain how much each individual takes from the family pool and how the total is allocated.

The Application of the Least Squares Principle

The method involves assigning constants to represent the consumption

of each type into which individuals have been classified, and then estimating the constants by minimising a certain sum of squares. It has already been described by Mr. Quenouille,[5] so it is unnecessary to give more than a bare outline. The actual categories into which individuals are classified may be chosen as convenient. The main illustrations in this paper use the following, with the constants assigned to each on the left:

a Child age 0–6 years
b Child age 7–14 years
c Male adolescent age 15–19 years
d Female adolescent age 15–19 years
g First male and female adults in family
e Each extra male adult
f Each extra female adult

Subsequently the categories are denoted by roman letters corresponding to the above italic letters. In this paper these constants relate to the cost of feeding individuals and not to consumption, because the data available gave the total expenditure on food. The methods are perfectly general and can be applied wherever it is required to analyse family totals into individual components.

With the constants defined above, the expenditure of a family of n members consisting of n_a of the first category, n_b of the second, and so on, may be represented by:

$$x = an_a + bn_b + cn_c + dn_d + en_e + fn_f + gn_g \qquad \ldots.(1)$$

Values of the constants can be estimated by minimising the quantity $\Sigma w_m(x_m' - x_m)^2$, where x_m' and x_m are the actual and estimated expenditures of the mth family. In this article two distinct sets of solutions are given. In one no difference is made between the weighting of different families, so that $w_m = 1$ for all families. In the other set $w_m = 1/n$, so that each family is given a weight inversely proportional to its size.

Some evidence suggesting that inverse weighting is appropriate is presented in the next section. It has, moreover, an added advantage to recommend it. In general, the derivation of the constants in equation (1) by least squares will not provide values which, when applied to a household of average composition, will yield the average expenditure. This relationship could be ensured by treating it as a linear constraint which the final solution must satisfy. The same result can be achieved automatically by weighting each household inversely by size. This may be readily proved by adding together the normal equations obtained for the weighted solution, remembering that the normal equation corresponding to n_g has to be added in twice.

It is of interest to note that Dr. Deming[6] gives the solution of the simple case when only one independent variable is involved, namely that if the equation to be fitted by least squares is $y = kx$ and the

variance of y is proportional to x, then $k = \Sigma y/\Sigma x$. The present case is a generalisation of this formula, in which the variance of the dependent variable is proportional to the sum of the independent variables.

The main data used to illustrate the method were obtained as a by-product of two surveys made for the Central Statistical Office in April and September 1949. Altogether a sample of 2391 housewives gave an estimate of what they usually spent on food in a week, with the results shown in Table I. No attempt was made to get itemised accounts, but this simple and direct approach gives sufficiently accurate data for our purpose although some non-food items such as matches and polishes may be included in some cases. Out of an expenditure of just under £3 per household about 2s. will be on items other than food. The group of 'old people' has been separated because it contains practically no children, and the expenditure and nutritional requirements of old people are lower.

TABLE I

Domestic expenditure on food: means and variances for households of different sizes

Number in Household (n)	Number of Households	Weekly Expenditure (shillings)	
		Mean per Person	Mean Square†/ No. in Household (s^2/n)
Main Sample			
1	118	20·3	84·1
2	490	22·0	137·9
3	639	18·9	98·5
4	465	16·3	89·0
5	224	14·3	119·6
6	88	14·5	125·9
7	51	14·0	121·4
8 or more	35	12·7	160·8
All households	2,110	17·2	—
*Old People**			
1	149	16·4	47·4
2	111	16·5	79·0
3	20	14·8	102·5
5	1	21·6	—
	281	16·3	—

* The head of the household in this group was 65 or over if male, or 60 or over if female.
† Mean square (s^2) of total household expenditure.

The point of immediate interest is the change in the variance (s^2) with size of household. The distribution is heteroscedastic and the variance increases with the number of persons in the household, as indicated by the behaviour of s^2/n in the last column of Table I. The average value of s^2/n in the main sample, obtained by weighting by the number of degrees of freedom in each group, is 109·7; the variation from this is not great except in the small group with 8 or more persons in the household. It is this relative constancy of s^2/n which is the justification for weighting families inversely according to their size. On the

other hand, one cannot regard the variance as coming from the same population, as can be seen by an application of Professor Bartlett's μ test for homogeneity of variances.

A similar relation between variance and size of household was observed in the Ministry of Food data referred to below in the last section, both for total expenditure and consumption and also for individual quantities. Weighting households inversely according to the size will therefore tend to ensure homoscedasticity, but it is unlikely that the resulting distribution will be normal. Fortunately, lack of normality does not invalidate the application of a least squares procedure, although it may affect the application of tests of significance.

Application of Various Least Squares Solutions to the Data

A number of different solutions were tried out in order to test the effect of using the weighted form as compared with the unweighted form, and of including or omitting certain variables. The results, that is the values of a, b, c, etc., are shown for a selection of these solutions in Table II. The computation of the coefficients of the normal equations from which the values of a, b, c, etc., were found presents no problems. The sums and cross-products can be readily formed by using Hollerith punched cards and the technique of progressive digiting. This is an extremely simple operation in the case of the unweighted products, since each variable can take only the integers 1, 2, 3, etc., or zero; and the highest value is usually below 9; in particular, n_g may be only 0 or 1.

TABLE II

Estimates of individual expenditure from different solutions, shillings per week

Type of Individual	Number	Whole Sample				Old People
		A	B	C	D	E
(a) Children 0–6 years	1,057	7·3	7·9	8·4	9·4	—
(b) Children 7–14 years	846(2)	9·1	9·4	10·1	10·8	—
(c) Boys 15–20 years	225	14·5	14·1	14·1	14·2	—
(d) Girls 15–20 years	269	11·0	14·7	15·9	14·8	—
(e) Men other than (g)	473(34)	15·7	16·5	16·8	22·2	14·6
(f) Women other than (g) ..	938(178)	9·1	13·3	15·4	18·0	15·8
(g) Man and woman	1,909(111)	32·6	43·7	43·2	—	34·0
Constant	—	12·8	—	—	—	—
		Standard error of estimates				
(a) Children 0–6 years		0·46	0·53	0·60	0·58	—
(b) Children 7–14 years		0·47	0·56	0·67	0·66	—
(c) Boys 15–20 years		0·68	1·21	1·42	1·41	—
(d) Girls 15–20 years		0·65	1·07	1·28	1·27	—
(e) Men other than (g)		0·56	0·80	0·80	0·49	3·77
(f) Women other than (g) ..		0·52	0·48	0·41	0·40	1·56
(g) Man and woman		0·71	0·58	0·53	—	2·85

NOTES. The whole sample includes old people, the number of whom is given in brackets in the first column of figures. A and B are solutions obtained without and C, D, and E with weights. Solution A is based on the usual regression technique and the equation has a constant term. In obtaining solution D all men and all women were treated as separate variables.

There is little difference between the values given by the first four solutions in Table II. These all contain the sample of 'Old People' and, although the values of this group when taken separately differ from the main solutions, the exclusion of this sub-group from the sample will have little effect on the results. This point was confirmed by repeating the calculations for solution C but with the 'Old People' households excluded; the results were almost identical with those for C.

Solution A naturally produces results for adults different from most of the other solutions owing to the presence of the constant term. The low values of f and g may be explained by most households (1909 out of 2391) containing at least one man and one woman, while of the 267 single-person households most are women (219) living on their own. Consequently much of the expenditure that in other solutions has been attributed to the two types (f) and (g) has been attributed to the constant. Likewise, the two adult values in solution D differ somewhat from those for other solutions because all except D treat the first man and woman in the household as one unit instead of treating all men and all women separately and alike.

In addition to the direct comparison of the estimates and of the precision with which they have been estimated, a further test of the solutions can be made by calculating the estimated expenditure for households of different sizes and comparing them with the corresponding mean expenditures for the sample as a whole. The actual means show that after households of two, expenditure declines with household size. This relation was reproduced by the calculated values obtained from all the solutions except A, which gave much too high a value for single-person households and rather too high a value for two-person households. A part of the decline in expenditure with size of household is therefore explained by the increasing proportions of children. This does not account for all the fall in expenditure, and some must be caused by a fall in the standard of living as household size increases.

From this examination one can conclude that although there are theoretical reasons for preferring the weighted solutions of type C, there seems to be no practical gain over solution B. The latter is the simplest to compute, so it is probably the best one to use. It is worth noting that for many purposes solution D gives satisfactory results and, as one less variable is involved, the amount of computing is smaller.

Finally, in spite of the theoretical objections to solution A, it gives values surprisingly close to those from other methods. Firstly, because it provides a closer fit to the observations it gives more precise determinations except for the term g referring to the first man plus the first woman in the household. As will be seen from the lower panel of Table II, this estimate for g has a higher standard error largely because much of the expenditure attributed by other solutions to this variable is here given to the constant term. Secondly, solution A has given by chance a more acceptable relationship between c and d, making the former the greater value. This relationship would be expected from

considerations of nutritional requirements and is satisfied by the results of the subtraction method described later and by the Ministry of Food data. No explanation can be given as to why the other solutions fail to conform, but in all these cases the difference between *c* and *d* is insignificant, the standard error of the difference being about two shillings; for solution A, however, it is no more than one shilling.

The Subtraction Method

Although the subtraction method of estimating individual expenditures is inefficient, since it ignores many households who do not fit into one of the standard compositions used in the comparisons, it has been used on a number of occasions and the following brief remarks may be of interest.

Table III shows the distribution of the 2110 households in the main

TABLE III

Numbers of households of various types and average weekly outlay per household in shillings

Household Type	No Adults except Man and Wife*		All Other†
(a) (b) (c) (d)	No.	Outlay	No.
No children			
0 0 0 0	387	45·5	446
One child			
1 0 0 0	224	54·5	62
0 1 0 0	123	57·7	43
0 0 1 0	37	64·0	32
0 0 0 1	34	63·4	45
Two children			
1 1 0 0	94	62·4	17
1 0 1 0	6	56·5	5
1 0 0 1	5	68·5	14
0 1 1 0	13	66·3	6
0 1 0 1	13	62·0	13
0 0 1 1	6	81·1	8
2 0 0 0	115	61·1	19
0 2 0 0	38	66·3	17
0 0 2 0	4	69·0	3
Three children			
1 1 1 0	6	72·7	2
1 1 0 1	2	67·6	1
1 0 1 1	3	75·1	—
0 1 1 1	3	76·5	4
1 2 0 0	31	72·0	6
1 0 2 0	1	110·0	1
1 0 0 2	1	126·1	2
0 1 2 0	1	69·6	1
0 1 0 2	1	80·0	2
0 0 1 2	1	80·3	2
2 1 0 0	25	70·8	6
2 0 1 0	4	53·1	—
2 0 0 1	1	31·0	5
0 2 1 0	7	70·8	5
0 2 0 1	7	71·7	4
0 0 2 1	1	69·7	—
3 0 0 0	23	58·3	6
0 3 0 0	9	73·1	4
Four children			
1 1 1 1	2	104·3	—
1 2 1 0	4	87·0	—
0 1 1 2	1	66·3	2
2 1 1 0	2	76·2	1
2 1 0 1	1	81·5	1
0 2 1 1	3	82·8	3
2 2 0 0	4	77·2	—
1 3 0 0	8	60·6	2
3 1 0 0	2	68·3	3
4 0 0 0	3	76·0	—
0 4 0 0	1	55·2	1
Five children			
1 2 2 0	1	109·8	—
1 2 0 2	1	97·1	—
2 1 0 2	2	79·5	—
2 2 1 0	2	116·7	—
2 2 0 1	2	80·2	3
0 2 2 1	1	140·0	—
1 3 0 1	1	146·5	—
2 3 0 0	3	89·8	—
3 2 0 0	2	96·0	—
0 5 0 0	1	114·7	—
Six children			
1 3 1 1	2	112·9	—
1 3 0 2	1	131·3	—
2 4 0 0	1	87·0	—
4 2 0 0	1	100·0	—
3 3 0 0	1	79·7	—
0 3 0 3	1	80·0	—
1 5 0 0	1	101·0	—
Seven or more children			
1 4 1 1	1	110·0	—
3 4 0 0	1	84·9	—
1 3 2 2	1	75·3	—
2 5 1 1	1	125·0	—
6 3 3 0	1	125·7	—
	1,288	—	797‡

* Households containing one man and one woman as the only adults, together with the number of children of various age groups specified in first and fifth columns where the groups are:

(a) 0–6 years
(b) 7–14 years
(c) 15–19 years (boys)
(d) 15–19 years (girls).

† All households containing either one adult or more than two.

‡ In addition to the 797 households listed, there were 25 not containing exactly one man and one woman and not with a structure falling into one of the 65 listed types, making a total of 822 households.

sample according to the composition of the household as defined by the number of persons under 20 in the four age-sex groups (a), (b), (c), and (d), defined in a footnote to the table. Even taking the 1288 in which the only adults were one man and woman, the definition of the composition in terms of these four broad age-sex groups has led to 65 different cases, a number which would be increased if the sample were larger. The figures in the fourth and eighth columns are the numbers of households ignored when the estimates are confined to those households containing a man and woman as the only adults.

The subtraction method proceeds by comparing families of recognisable and simple types and for this reason it is usually confined to households containing a man and woman as the only adults. All families containing fewer or more than this number are discarded; here this is 822 or almost 40 per cent. The expenditure on a child 0–6 years, for example, is estimated by comparing families of a man, woman, and one such child, with families of a man and woman only. In the notation of Table III this implies subtracting 45·5 shillings, the average cost of the 387 families of type (0000), from 54·5 shillings, the average cost of the 224 families of type (1000). This is the first comparison listed in Table IV.

TABLE IV

Expenditure on each child estimated by subtraction method, shillings per week

Child Type	Family Types of Table III in Comparison	Expenditure on Additional Child	Standard Error of Expenditure
(a) 0–6 years	(1000)–(0000)	9·0	1·43
(b) 7–14 years ..	(0100)–(0000)	12·2	1·79
(c) Boys 15–19 years	(0010)–(0000)	18·5	3·08
(d) Girls 15–19 years	(0001)–(0000)	17·9	3·20

Some idea of the loss of information may be seen from the proportion of the sample utilised in making each estimate, but an exact method is to compare the reliability of each estimate obtained by subtraction with that from least squares. The standard errors for the latter estimates are given in Table II and those for the subtraction estimates can be calculated from:

$$\left(\frac{2v}{m_2}+\frac{3v}{m_3}\right)^{\frac{1}{2}}$$

where v is the average value of s^2/n as defined earlier and which for the main sample in Table I was 109·7; m_2 and m_3 are the number of families in the two groups used in the comparisons, m_2 being the number without children and m_3 the number with the selected child. The justification of this formula depends on the correctness of the assertions made above that the variance varies inversely with size of family.

The magnitude of the loss in information caused by using the subtraction method as against the least squares solutions is quite striking

as soon as Tables IV and II are compared. In the former the standard errors are about three times those in the latter. Some improvement might be possible by making further comparisons from the data of Table III. For example, the comparison (1100)–(0100) gives an estimate of 4·7 shillings for a child of type (*a*). Additional comparisons of this kind are difficult to use, since there appears to be no satisfactory method of combining them. They could be averaged together by taking account of their standard errors calculated as in Table IV, but strictly speaking this would be incorrect, since there is correlation between many of the estimates.

If further comparisons are made by using the data of Table III it will be found that the estimates based on families with two or more children are lower than those in Table IV, which are confined to families with one child. This might be regarded as evidence of decreasing food expenditure on additional children, but the inference must be qualified by taking account of the low reliability of the determinations of the values for children from families with two or more children. Some of this lower expenditure probably reflects the equivalently lower standard of living of those with larger families: lower because there are more dependents to be covered by the same income. As explained in the introduction, this aspect can only be dealt with properly if it is possible to find some method of assessing the standard of living of the family, a subject which is outside the scope of this paper.

Estimates from Ministry of Food Survey

Table V gives results derived by the author while working in the

TABLE V

Individual consumption in 1944, *Ministry of Food survey*

	Daily Calorie Intake	Weekly Expenditure	Weekly Bread* Consumption April–June only	Weekly Potato Consumption Oct.–Dec. only
		shillings	ounces	ounces
Children 1–3	1,823	7·1	49	57
Children 4–6	1,990	7·4	67	62
Children 7–9	2,266	8·5	74	67
Children 10–12 ..	2,423	8·5	106	75
Boys 13–15	2,639	9·3	123	82
Boys 16–20	3,059	12·4	137	109
Girls 13–15	2,271	8·3	80	75
Girls 16–20	2,334	9·9	98	86
Men	2,680	11·9	96	87
Women	2,280	11·0	75	61
Expectant mothers ..	2,634	12·3	83	86
Mother with child under 1 year	3,879	17·4	95	105
Average consumption or expenditure	2,387	10·6	82	72
Number of households	7,623	7,623	2,087	1,760
Number of persons ..	26,015	26,015	6,984	6,131

* Includes flour, cakes, buns, and scones.

Ministry of Food on the inquiry[7] in 1945–6. At the time the calculations were made they gave estimates of individual food consumption not otherwise available, which were needed to check ration scales and therefore were of direct administrative importance.

The least squares method used was one of the simplest types of solution in that the unweighted form was used considering men and women as separate variables and without removing old people from the sample. From the comments on Table II it would appear that similar results would have been obtained even if a more sophisticated treatment had been adopted. The estimate of the combined consumption of a mother and a child under one year is on the high side, but this value is probably not very reliable, since there were few families with infants.

The estimates of food expenditure on each individual may be compared with those given in the present paper. Although average expenditure in the 1949 data was about 60 per cent. above the 1944 level, the estimates of individual cost are by no means all increased in the same ratio. At one extreme the cost of a child 0–6 years appears to have increased by no more than 10 per cent., while at the other the cost of feeding a man and woman together appears to have gone up by 90 per cent. This is not an exact comparison, since the two sets of data do not refer to comparable households and were not obtained in the same way. Nevertheless, it does suggest that the dispersion in the expenditure on different types of persons may have increased.

The author is indebted to the Central Statistical Office and the Ministry of Food for permission to use the data on which this paper is based, and to the staff of the Social Survey who carried out many of the calculations.

REFERENCES

[1] Clements, F. W. (1940). 'A Family Coefficient Scale developed from the Australian Nutrition Survey.' *J. Hygiene*, **40**, 681.

[2] Widdowson, E. M. (1947). *A Study of Individual Children's Diets*. MRC Special Report, Series No. 257.

[3] Nicholson, J. L. (1949). 'Variations in Working Class Expenditure.' *J. R. Statist. Soc.* A, **112**, 359.

[4] Houthakker, H. S. (1952). 'The Econometrics of Family Budgets.' *J. R. Statist. Soc.* A, **115**, 1.

[5] Quenouille, M. H. (1950). 'An Application of Least Squares to Family Diet Surveys.' *Econometrica*, **18**, 27.

[6] Deming, W. E. (1943). *Statistical Adjustment of Data*, p. 31. Wiley, New York.

[7] Ministry of Food (1951). *The Urban Working Class Household Diet, 1940 to 1950*. HMSO, London.

COMPOSITE MEAN SQUARES AND THEIR DEGREES OF FREEDOM

LEROY P. V. JOHNSON and ERNEST S. KEEPING
University of Alberta, Edmonton, Alberta, Canada

This article calls attention to a relatively little-known method for dealing with a situation that sometimes arises when treating experimental results by the analysis of variance. It is intended for readers who are acquainted with the standard methods of analysis.

WHEN three or more main effects are involved in an analysis of variance three or more interaction effects may occur as sources of variation. This may lead to circumstances in which no term in the analysis is a valid error for testing the significance of a given main effect. A way in which this may come about may be shown by examination of the components of variance in the following hypothetical three-criteria analysis.

Source of Variation	D.F.	Expectation of Mean Squares	Mean Squares
Varieties (V) ..	4	$\sigma^2 + 3\sigma_{VL}^2 + 4\sigma_{VY}^2 + 12\sigma_V^2$	$M_V = 20$
Locations (L) ..	3	$\sigma^2 + 3\sigma_{VL}^2 + 5\sigma_{LY}^2 + 15\sigma_L^2$	$M_L = 15$
Years (Y)	2	$\sigma^2 + 4\sigma_{VY}^2 + 5\sigma_{LY}^2 + 20\sigma_Y^2$	$M_Y = 10$
V × L	12	$\sigma^2 + 3\sigma_{VL}^2$	$M_{VL} = 5^*$
V × Y	8	$\sigma^2 + 4\sigma_{VY}^2$	$M_{VY} = 6^*$
L × Y	6	$\sigma^2 + 5\sigma_{LY}^2$	$M_{LY} = 3$
V × L × Y (E) ..	24	σ^2	$M_E = 2$

* Significant above the 5 per cent. level.

It is assumed that the particular varieties, locations, and years are to be regarded as random samples from very large populations. This is the common assumption in technical agriculture. The conclusions to be drawn from the experiment are intended to be good for *all* varieties, locations, and years.

The second-order interaction constitutes a valid error for testing the significance of any first-order interaction, two of which are significant. In this case M_{VL} is a valid error for testing the significance of M_L, since σ_{LY}^2 may be assumed to be equal to zero. Similarly, M_{VY} is a valid error for testing M_Y. But none of the interactions is valid as an error for testing the significance of M_V, since one or more significant components in this mean square will be disregarded.

In an effort to overcome such difficulties a method has been given by Snedecor[5] and Johnson[1] which involves the computing of a valid error term as a composite of appropriate components derived from (usually) three mean squares. For example, in the present case the valid error for testing M_V (i.e. the null hypothesis, $\sigma_V^2 = 0$) must have the constitution $\sigma^2 + 3\sigma_{VL}^2 + 4\sigma_{VY}^2$ which results from $M_{VL} + M_{VY} - M_E$. This method, however, does not permit computation of the degrees of freedom associated with the composite mean square. This difficulty is overcome in the following method, which involves a composite mean square for the main effect as well as a composite mean square for the error. This method has been given by Satterthwaite[4] and is described in Kenney & Keeping's text-book.[2] It does not appear, however, to be generally known.

Assuming that we wish to test the significance of M_V in the above example, the variance ratio will be $F = (M_V + M_E)/(M_{VL} + M_{VY})$, which gives a test of the null hypothesis $\sigma_V^2 = 0$, since the expectation of the numerator is $2\sigma^2 + 3\sigma_{VL}^2 + 4\sigma_{VY}^2 + 12\sigma_V^2$ and that of the denominator is $2\sigma^2 + 3\sigma_{VL}^2 + 4\sigma_{VY}^2$. The numerator and denominator are not, however, strictly distributed as χ^2 so that the ratio is not strictly distributed as F. The method to be described is admittedly approximate, but exact values of the degrees of freedom can scarcely be regarded as important when the significance levels are themselves arbitrary. The hope is that the method will enable users to obtain a rough idea of the significance of the main effects in circumstances in which the ordinary analysis of variance is inapplicable. Even the standard analysis makes several assumptions, such as independence, normality, and constancy of variance of the observations, which can only be considered in general as rough approximations to the truth.

If a quantity q is distributed as χ^2 with n degrees of freedom, the expectation and variance of q are given by

$$\mathrm{E}(q) = n, \qquad V(q) = 2n \qquad \ldots.(1)$$

On the assumptions usual in analysis of variance the constituent sums of squares, divided by the appropriate variances, are each distributed as χ^2 with the corresponding number of degrees of freedom. From the table above and equation (1):

$$\left.\begin{aligned} \mathrm{E}[4M_V/(\sigma^2 + 3\sigma_{VL}^2 + 4\sigma_{VY}^2 + 12\sigma_V^2)] &= 4 \\ \mathrm{V}[4M_V/(\sigma^2 + 3\sigma_{VL}^2 + 4\sigma_{VY}^2 + 12\sigma_V^2)] &= 8 \end{aligned}\right\} \qquad \ldots.(2)$$

Therefore $\mathrm{E}(M_V) = \sigma^2 + 3\sigma_{VL}^2 + 4\sigma_{VY}^2 + 12\sigma_V^2$

and $\mathrm{V}(M_V) = 2(\sigma^2 + 3\sigma_{VL}^2 + 4\sigma_{VY}^2 + 12\sigma_V^2)^2/4$

since the variance of cx is c^2 times the variance of x. It follows that

$$\mathrm{V}(M_V) = 2\mathrm{E}^2(M_V)/4 \qquad \ldots.(3)$$

and in the same way $\mathrm{V}(M_E) = 2\mathrm{E}^2(M_E)/24 \qquad \ldots.(4)$

We now assume that the distribution of $M_V + M_E$ is of the same form as the distributions of M_V and of M_E but with a number of degrees of freedom n_1 to be determined. That is to say, if $\mathcal{N} = M_V + M_E$ we assume that

$$V(\mathcal{N}) = 2E^2(\mathcal{N})/n_1 \quad \ldots\ldots(5)$$

Since $E(\mathcal{N}) = E(M_V) + E(M_E)$ and $V(\mathcal{N}) = V(M_V) + V(M_E)$ it follows from (3) and (4) that

$$2E^2(M_V)/4 + 2E^2(M_E)/24 = 2[E(M_V) + (M_E)]^2/n_1 \quad \ldots\ldots(6)$$

The expectations are unknown but can be replaced approximately by their estimates, M_V and M_E. We then have an equation to determine n_1. Thus, in the example given, $2(20)^2/4 + 2(2)^2/24 = 2(22)^2/n_1$, whence $n_1 = 4{\cdot}83$ (5 to the nearest integer).

Similarly, if $D = M_{VL} + M_{VY}$:

$$V(D) = 2E^2(D)/n_2 \quad \ldots\ldots(7)$$

so that

$$2E^2(M_{VL})/12 + 2E^2(M_{VY})/8 = 2[E(M_{VL}) + E(M_{VY})]^2/n_2 \ldots\ldots(8)$$

whence $n_2 = 18{\cdot}38$ (18 to the nearest integer). If we now put

$$F = \mathcal{N}/D = (M_V + M_E)/(M_{VL} + M_{VY}) \quad \ldots\ldots(9)$$

we know that $E(\mathcal{N}) = E(D)$ on the null hypothesis that $\sigma_V{}^2 = 0$.

If also $\mathcal{N}$ and D have the χ^2 type of distribution assumed above, then F is distributed as Snedecor's F and the usual significance test is valid. Thus, in the example, $F = 22/11 = 2$, with 5 and 18 degrees of freedom. The 5% value is 2·77, so that the variety effect is not significant.

It will be seen that there are three approximations in this procedure: (1) the replacement of the actual distributions of M_V, M_E, and $M_{VL} + M_{VY}$ by χ^2 distributions; (2) the use of the F distribution; and (3) the replacement of expected values by their estimates from the experiment.

(1) As regards the first approximation, it is known that if q_1 and q_2 are independently distributed as χ^2, and if $q = aq_1 + bq_2$, then q is also so distributed if $a = b$ or if a or $b = 0$.

(2) Welch[6] has used a Pearson Type 3 curve, with the same mean and variance as q, to approximate the distribution in other cases, and has shown that the fit in the examples studied is reasonably good. The exact distribution of q can be obtained, but not in any very convenient form. The density function is

$$f(q) = Cq^{m+n-1}e^{-q/2a}\int_0^1 (1-x)^{m-1}x^{n-1}e^{-qx(a-b)/2ab}dx \quad \ldots\ldots(10)$$

where and

$$\left.\begin{aligned} C^{-1} &= 2^{m+n}\Gamma(m)\,\Gamma(n)a^m b^n \\ m &= n_1/2,\ n = n_2/2 \end{aligned}\right\} \quad \ldots\ldots(11)$$

In the example above, $4M_V/E(M_V)$ is distributed as χ_4^2 and $24M_E/E(M_E)$ as χ_{24}^2 so that $\mathcal{N}$ is distributed as $a\chi_4^2 + b\chi_{24}^2$, with $a = E(M_V)/4$ and $b = E(M_E)/24$. If we use the estimates of M_V and M_E for the expectations, $\mathcal{N}$ is distributed as $5\chi_4^2 + \frac{1}{12}\chi_{24}^2$. Here $m = 2$ and $n = 12$. For these values the integral in (10) can be evaluated in terms of the incomplete gamma function which has been tabulated by Karl Pearson.[3] In terms of the tabulated function I (u, p):

$$f(q) = C_1 e^{-0.1q}[5.9q\,I(5.9q/\sqrt{12}, 11) - 12\,I(5.9q/\sqrt{13}, 12)] \quad (12)$$

where $C_1 = 0.002074$. The assumed equivalent distribution is that of $\chi_5^2[E(M_E + M_V)]/5$, or roughly $4.4\chi_5^2$, for which the density function is

$$g(q) = C_2 q^{3/2} e^{-5q/44},\ C_2 = 0.0032746 \quad \ldots.(13)$$

The two distributions are of the same general form, particularly for the larger values of q. As q increases, $f(q)$ approximates to the form

$$f(q) = C_1 e^{-0.1q}(5.9q - 12) \quad \ldots.(14)$$

The 5 per cent. point for (12) is given by $q = 49.5$ and that for (13) by 48.7. These are the values of q such that the area under the curves beyond q is 0.05. It appears that even with a and b as different as 5 and 1/12, the χ^2 approximation is tolerably good. If so, the F distribution may reasonably be expected to give approximately correct 1 per cent. and 5 per cent. points.

(3) There remains the error involved in replacing expectations by estimates. This, of course, is unpredictable. It is well known that estimates, particularly with few degrees of freedom, are subject to considerable variation. Thus it is seen from equation (3) that the standard deviation of M_V may be estimated as $20/\sqrt{2} = 14$, and similarly an estimate of the standard deviation of M_E is 0.6. The value of M_V, certainly, is not very reliable. This use of a single estimate of a variable quantity is often, however, unavoidable in experimental work.

REFERENCES

[1] Johnson, L. P. V. (1950). *An Introduction to Applied Biometrics.* Burgess, Minneapolis. (See pp. 114–15.)

[2] Kenney, J. F., and Keeping, E. S. (1951). *Mathematics of Statistics*, Part II. Van Nostrand, New York. (See pp. 262–3.)

[3] Pearson, K. (1922). *Tables of the Incomplete Gamma Function.* HMSO, London.

[4] Satterthwaite, F. E. (1946). *Biometrics*, **2**, 110–14.

[5] Snedecor, G. W. (1946). *Statistical Methods* (4th edition). Iowa State College Press, Ames, Iowa. (See pp. 308–9.)

[6] Welch, B. L. (1937). *Biometrika*, **29**, 350–62.

QUESTIONS AND ANSWERS

Readers are invited to submit questions or to deal shortly with points of general interest by supplying both the question and the answer. The first three of the following questions concern less-known sources of data.

Question

Can you inform me of any publication which gives the number of domestic electricity consumers in each town in Great Britain ?

Answer

No official body publishes the information in the form required. The *Electricity Supply Handbook* (published by the Electrical Times, price 5s.) shows, however, for the majority of the fourteen electricity Areas, the number of consumers, population, and acreage of each electricity district within those Areas. Unfortunately the districts are not co-terminous with the local government administrative areas, but the population figures published in the *Handbook* may be of some assistance in defining coverage of electricity districts in terms of administrative areas.

E. SHANKLEMAN

Question

For a study I am making of the comparative standards of living in Great Britain and the United States I am keen to obtain statistics of: (*a*) population and land area of both countries; (*b*) the number of private motor-cars in use; (*c*) the number of telephones in use; (*d*) the number of radio sets in use; (*e*) mileage of railways; (*f*) mileage of roads; and (*g*) electricity capacity.

Answer

The data sought are available in the following publications: (*a*) United Nations Statistical Year Books; (*b*) *The Motor Industry of Great Britain*, 1950, published by The Society of Motor Manufacturers, 148 Piccadilly, London, W.1; (*c*), (*d*), and (*e*) *The World Almanac*, published by The New World-Telegram and the Sun; (*f*) *World Road Statistics*, published by The International Road Federation Ltd., 18 South Street, London, W.1.; and (*g*) *Statistical Year Book*, published by The World Power Conference, Trafalgar Buildings, Charing Cross, London, S.W.1.

E. SHANKLEMAN

Question

Where can I find data on the incidence of deafness among people in Great Britain ?

Answer

The only source where figures are available on this subject is the

Social Survey report *The Prevalence of Deafness in England, Scotland and Wales* published in 1948. This gives *inter alia* a regression of the prevalence of deafness on age, with analyses of sex and occupation.

E. SHANKLEMAN

Question

Reference is frequently made in population statistics to 'conurbations.' Could you please define them; where is detailed information available as to their composition?

Answer

Both requirements are met by the recently published Part I of the volume of *One per cent Sample Tables* prepared from the 1951 Census enumerations by the Registrars-General of England and Wales, and Scotland.

'Conurbations' may be defined as enclosing 'continuously urbanised areas surrounding large population centres which are to a greater or lesser extent focal points of economic and social activity.' The following seven conurbations are identified; they are to be used in future in the presentation of official statistics generally. E. SHANKLEMAN

	1951 Census Population (*thousands*)	Percentage of Total Population in Great Briatin
Greater London	8,344	17·08
South-east Lancashire	2,421	4·96
West Midlands	2,237	4·58
Central Clydeside	1,758	3·60
West Yorkshire	1,692	3·46
Merseyside	1,382	2·83
Tyneside	835	1·71
Total Population in Conurbations ..	18,669	38·22

Question

The statistician often has to explain the result of an *F*-test in non-statistical terminology. If the result of such a test is that the variance of A is x^2 times that of B, should he say that 'A is x^2 times as variable as B' or 'A is x times as variable as B'?

Answer

Most technicians think of variability as a linear measurement, and accordingly, if the standard deviation of process A is x times that of process B, this would be understood to mean that A is x times more variable than B. But variability must not be confused with precision. It should always be made clear that if A is x times more variable than B, then B is x^2 times more precise than A; that is to say, in order to obtain the same precision in an average, x^2 times as many observations will be required with A than with B. O. L. DAVIES

NOTES AND COMMENTS

THERE was great satisfaction in statistical circles at the appearance in the last Birthday Honours List of the name of Professor Ronald Aylmer Fisher as having been awarded a knighthood. Sir Ronald Fisher is President of the Royal Statistical Society for the current session.

★ ★ ★

A joint conference of the Royal Statistical Society and the Manchester Statistical Society was held at Ashburne Hall, Manchester, during the week-end 19th–22nd September 1952. The papers were as follows: Regional Economic Statistics, with Special Reference to the North West Region: K. S. Lomax; The Use of the Electronic Computer in Business and Economic Statistics: B. B. Swann; The Development of the Cotton Industry and Trade: R. Robson; A Survey of Electricity Consumers: G. R. Reah and T. R. Smith; The Consumer as Forecaster: Henry Durant; Investigations into Productivity in Cotton Spinning: L. H. C. Tippett and P. D. Vincent; Rational Sampling in Industry: G. H. Jowett; A Regression Study in Steel Making: D. R. G. Davies; The Meaning and Practical Importance of Stochastic Processes: M. S. Bartlett; Non-Normality and Tests of Variances: G. E. P. Box; Experiments for comparing Variability in Chemical Analysis: D. H. Ward.

Visits of general interest were paid to the Shirley Institute, to the cotton spinning and weaving mills of Messrs. Ashton Bros. Ltd., Hyde, and to the docks of the Manchester Ship Canal; and the Manchester Universal Electronic Computer was demonstrated to parties that visited Messrs. Ferranti Ltd. at Moston and the University.

★ ★ ★

The electronic computing machine, ACE, at the National Physical Laboratory is now able to accept and perform work for industry at a charge depending on the size and nature of the problem. For example, the charge for solving linear simultaneous equations with n unknowns is $n^2/50$ pounds, with a minimum of £2. The maximum for n is at present 62. The punching of the coefficients on the cards (standard Hollerith cards are used) costs an additional 10 per cent. and the matrix solution, involving n different right-hand sides to the equations, a further 50 per cent. This matrix solution gives the Fisher c coefficients which are used in calculating the variances of the estimates of the coefficients. Modern developments in computing, of which this service provides an example, open up possibilities of investigations involving multiple regression analyses, that were previously regarded as impracticable on the grounds of cost.

★ ★ ★

Statisticians often have to present data to laymen and commonly use pictorial diagrams and graphs in the belief that they assist the reader to

comprehend the facts. Miss M. D. Vernon in two articles in the January and April 1952 numbers of *Occupational Psychology* describes some experimental results that provide a sobering corrective to an over-enthusiastic acceptance of this belief. In the first experiments statistical material was presented in several forms to members of the armed forces, to university students, and to grammar school pupils, and it was found that only the better-educated and more intelligent subjects could learn much from diagrams and graphs and that even for these the material had to be fairly simple. In the second experiments diagrams and graphs were used to illustrate a written text presented to boys and girls from the sixth forms of grammar schools. Miss Vernon's conclusions are well stated in her own words, which we quote by permission:

'. . . a written argument will not be made easier to understand and remember by presenting factual data graphically except perhaps in certain special circumstances. The most important condition is that the readers shall have sufficient intelligence and general education to be able to handle graphical material efficiently. And, judging from the experiments carried out in this investigation, this facility does not occur very widely except in those who have had a university education, or possess its equivalent in intelligence and ability.

'This conclusion may be qualified by saying that understanding may be more widespread with a simple text accompanied by simple charts or graphs. . . . [But] they are valuable only in so far as they can be perceived to corroborate or extend the facts upon which the argument of the text is based.'

★ ★ ★

Attention was directed to the Russian attitude to statistics as a science by Dr. Stuart A. Rice last December in his presidential address to the Conference of the International Statistical Institute held in New Delhi. Dr. Rice's address is printed in the June–July 1952 issue of *The American Statistician*, together with a translation of a report of the Methodology Conference of the Central Statistical Administration of the USSR held in February 1950. Russia does not participate in the activities of the ISI and Dr. Rice based his address on reports of the Russian Conference. It appears that Soviet statistics is concerned with social and economic topics and it is fundamental that the science 'serves as a weapon in the struggle for building a Communist society.' With this premise it is neither surprising nor unreasonable that Soviet statistics should be 'a social, class, and Party science' and that its theoretical basis should be 'dialectical and historical materialism and Marxist-Leninist political economy.' Statistical methodology based on probability and the 'law of large numbers' exists but, because chance does not operate in the social and economic field, it is wrong, in the Russian view, to include such methodology in statistics. Much of the practice within the somewhat limited field of Soviet statistics seems, however, to be the same as in the corresponding parts of 'bourgeois statistics.'

INDUSTRIAL APPLICATIONS SECTION OF THE ROYAL STATISTICAL SOCIETY

List of Meetings

BIRMINGHAM AND DISTRICT GROUP

30th January, 1952. Problems of Quality Control in a Rubber Factory: M. G. Peakman

27th February, 1952. Short Cut Analysis of Variance based on Range: H. O. Hartley

26th March, 1952. Congestion Problems in Industry: J. Murdoch

LONDON GROUP

1st February, 1952. The Study of Unbalanced Data from a Process including Several Recurring Changes: Joan Keen and D. J. Page

7th March, 1952. Some Applications of Statistical Methods in Concrete Research: A. Battersby

4th April, 1952. Stochastic Problems of Storage Systems: H. Herne and D. G. Nickols

2nd May, 1952. The Extraction of Data from Old Records: W. A. Pridmore

NORTH EASTERN GROUP

16th April, 1952. Significance Tests in Action: J. M. Runcie

MERSEYSIDE GROUP

25th February, 1952. A Control Chart for the Percentage of a Chemical Ingredient: E. F. Livesey

11th March, 1952. Answers to Members' Statistical Problems: B. P. Dudding

2nd May, 1952. Statistically Designed Experiments in the Chemical Industry: G. P. Box

SHEFFIELD GROUP

27th March, 1952. Computational Short Cuts and Calculating Machines: H. O. Hartley

24th April, 1952. The Accuracy of the Sampling of Coal: J. Hebden and G. H. Jowett

SOUTH WALES GROUP

23rd November, 1951. Statistical Methods for Production Engineers: E. A. G. Knowles

14th December, 1951. Quality of Information: L. T. Wilkins

25th January, 1952. Statistical Aids to Management: L. H. C. Tippett

22nd February, 1952. The Precision of Statistical Tests: D. G. Beech

25th April, 1952. A Discussion on Sequential Analysis: J. Parry Lewis

STUDY SECTION OF THE ROYAL STATISTICAL SOCIETY

Reports of Meetings

LONDON

9th April, 1952. The Problem of Estimating Housing Needs: M. J. Elsas

As Dr. Elsas was ill and unable to be present, his paper was read by Miss Solomon, Secretary of the Housing Centre in London. It was with deep regret that the Section learned soon after that Dr. Elsas had died on 18th April.

Dr. Elsas emphasised that his paper was concerned, not with housing demand but with needs, which could be classified according to the type of family: families without a dwelling of their own, families living in slums and who therefore need a new dwelling, and single-person families.

The lack of data made very difficult the problem of assessing needs. Important factors were: internal and external migration, changes in economic and social policy (social security and full employment had increased housing needs by causing numbers of marriages to increase), the policy of rent restriction, the proportion of vacant dwellings required to allow for labour mobility (Dr. Elsas suggested 4–5 per cent. was an appropriate figure), and the sharing of dwellings.

The Registrar-General's attempts to assess housing needs for the decade 1931–41 were reviewed. It had been estimated that the number of private families would increase by 917,000 and that 1,700,000 houses in all would be required. In fact, 2,559,000 houses were built; needs had increased more than was anticipated owing to improved social and economic conditions.

At the outbreak of the war the housing shortage was nearly overcome, but it was not possible to see then all the implications of full employment and social security. At the end of the war it was not clear whether the wartime increase in number of marriages would continue, and this made forecasting difficult. But it was clear that residential building was a major task and that the predominance of small families necessitated smaller dwellings on average than hitherto.

In the period 1891–1931 the number of families had increased by 37½ for every 100 marriages. On a similar basis Dr. Elsas estimated from the 4·6 million couples married between 1939 and 1950 that there were 2·3 million families for whom dwellings needed to be found. About 1·1 million new houses had been provided in that period thus leaving the housing shortage at about the same figure. To this, 0·9 million needed to be added for replacement of slums—about 7 per cent. of existing houses. Altogether this meant that 300,000 houses would be required annually for the next seven years.

Other evidence often quoted as to the size of housing needs included the extent of local authority housing lists. Dr. Elsas believed that there was a

certain amount of duplication in these lists and that they included families who had found suitable accommodation but not notified the authorities. Also, investigation in London and Cardiff had revealed that some 40 per cent. of applications could be defined as 'less urgent.' On a basis of a sample of 20 towns of all sizes Dr. Elsas had estimated that there were about 2·2 million families on waiting lists in England and Wales. Assuming 40 per cent. of these were less urgent, it seemed that about 1·3 million houses were urgently needed, a figure which agreed approximately with the estimate given earlier.

Dr. Elsas discussed the estimates of the future numbers of families, prepared in 1945 and published in the Report of the Economics Committee of the Royal Commission on Population. The Committee had seriously underestimated the increase in the number of families between 1945 and 1950.

11th June, 1952. Research on Radio Talks: Joseph Trenaman

Mr. Trenaman has been engaged for three years on research work into the educational effects of radio talks under the BBC's Further Education Experiment. A BBC engineer accompanied him and played over demonstrations of some of the talks which had been used in his experiments.

Mr. Trenaman explained the difficulty of assessing 'understanding,' which was something much deeper than the comprehension of the individual words or even the separate ideas in a radio talk. He himself felt that the use of easily comprehended words was less important than the structure of the ideas in the educational material, or the personality of the speaker, or, above all, the extent to which the listener was interested in the subject matter. Discussing tests which had been used, he quoted the answers of one listener who was able to reproduce many of the technical words in a talk, but failed completely to make sense of the speaker's general argument. It was found that even though a physicist was asked, in giving a talk, to address the general public and had taken great pains to make his material understandable, his talk could not be grasped by more than 10 per cent. of the total public. On the other hand, some of the educational talks were understandable by nearly 60 per cent. of the public—a satisfactory result considering the large proportions who read very little. Tests on adult students had shown that, although they were often anxious to listen for long periods, it was difficult for them to retain enough of a broadcast talk after 30 minutes to allow them to piece together what they had heard into a coherent whole.

Mr. Trenaman had applied indirect tests through educational institutes who had included in students' examinations some questions relating back to the broadcasts; and a comparison of the results of those who had heard a radio programme with those who had not showed that a great deal of the content of the programme is retained and used in new situations. Up to 90 per cent. of what was 'understood' immediately after listening was brought out subsequently by some groups. In comparing listeners' interest in an educational talk with their understanding, Mr. Trenaman had observed that the most enthusiastic were almost invariably those who had but barely understood the talk.

INDEX

accident records, 142
ACE, 208
agricultural research, 72
Allen, R. G. D., 106
Anson, A., 75
Aston, H., 72

Bailey, N. T. J., 149
Barnard, G. A., 141
Beech, D. G., 140
book reviews, 80, 148
Box, G. E. P., 141, 143
Bushell, F. D., 75

calibration of flowmeter, 118
Campbell, H., 142
cards, counting device, 139
Cauter, T., 75
Census of Distribution, 63, 146
chemical analysis of pyrethrum, 95
chemical industry, experiments in, 141, 143
Clark, D. A., 146
coal sampling, 179
coal technology, 72
computing equipment, 71
Congrès Internationale d'Organisations Scientifiques, 21
consumer panels, 78, 169
control chart technique, introduction and administrative details, 3, 15
conurbations, 169
conveyor belts, sampling from, 50, 179
Corlett, T., 34
Cost of Living Technical Committee, 106
Cox, P. R., 81, 82, 145
Crump, N., 77
Cunliffe, Stella V., 75

Davies, O. L., 148, 207
deafness statistics, 206
decision functions, 141
Desmond, D. J., 141, 142
Dobbie, W. R., 71
diagrams, usefulness, 209
Dudding, B. P., 3
Dyke, G. V., 95

Editorial, 2, 81
electric lamps, testing, 5
electricity statistics, 206
electronic computer, 208
Elsas, M. J., 211
epidemiology, 154
error curve, 144
expenditure of individuals, 192
experiments, design, 78
experiments in chemical industry, 141, 143
Eysenck, H. J., 45

factor analysis, uses and abuses, 45
'Facts from Figures,' review, 80
financial statistics, 77
financial statistics, sources, 60
Fisher, R. A., 208
food, expenditure on and consumption of, 192
foundry control, 144
Franklin, N. L., 143

Garwood, F., 145
gas flowmeter, calibration, 118
Gold, D., 143
Grace, A. D., 69
graphs, usefulness, 209
Greer, D., 63

Hebden, J., 179
Herdan, G., 69
Heyworth, G., 63
Hill, A. Bradford, 1
Hinsley, J. F., 144
history and statistics, 63
hospital statistics, 75
housing needs, estimation, 211
Hunter, F., 72

index numbers of retail prices, 106
Industrial Applications Section of the Royal Statistical Society, 65, 138, 140, 210
industrial health, 71
industry, statistical methods in, 3
inspection methods, 72, 144
International Management Congress, 1951, 21

Jenkins, W. G. A., 144
Jessop, W. N., 131
Johnson, L. P. V., 202
Jowett, G. H., 50, 62, 148, 179
jury qualification, use in sample design, 34

Keen, Joan, 67, 163
Keeping, E. S., 202
Kemsley, W. F. F., 192
Kendall, W. L., 146
Kenney, P. K., 78
Knowles, E. A. G., 140, 210
Komlosy, G. F., 144
Körner, S., 146

management, statistics and, 21, 143
Manchester Statistical Society, conference, 208
Mantell, J. E., 68
market research, 70, 75
market research, sampling, 60
mechanical recording, 71
mean squares, 202
medical genetics, 158
medical statistics, 149
medical treatment, 152
Moroney, M. J., 80

National Food Survey, 146

output, measurement of changes, 61

Page, D. J., 163
paint industry, 68
personnel management, 68
plastics, specification, 37
population estimates, 82, 145
population statistics, of USA and GB, 206
pot lives, 163
precision of statistical tests, 140
probability in statistics, 146
productivity measurement, 70, 142
pyrethrum, chemical analysis, 95

radio sets, statistics, 206
radio talks, research on, 212
railway mileage, 206
Ramm, B. T., 126
randomness, 147
regression analysis, application to calibration of flowmeter, 118
regression analysis, discussion of choice of lines, 119, 131
retail prices, index numbers, 106
retail sales, measurement, 126
Rice, S. A., 209
Rivett, B. H. P., 72, 144
road accident statistics, 145
road mileage, 206
Roseman, C., 70
Royal Commission on Population, 82
Royal Statistical Society, conference, 208
Runcie, J. M., 72

sales, measurement, 126
Salter, G., 60
sample design, use of jury qualification, 34
sampling for market research, 60
sampling from conveyor belts, 50, 179
Sears, G., 69
sequential sampling, application in steel works, 144
Shankleman, E., 75, 207
sieving process, 69
Silvey, R. J. E., 79
Simpson, E. P., 95
social class of towns, 27
socio-medical surveys, 149
Soviet statistics, 209
specification of plastics, 37
Spendley, W., 118
Squirrell, N. J., 70, 71, 142
'Statistical Methods for Chemists,' review, 148
Study Section of the Royal Statistical Society, 74, 145, 211
subjective judgment, 69
Swan, A. W., 21

Taylor, A. J. P., 63
'Teaching of Statistics in Schools,' 138
'Technological Applications of Statistics,' review, 148
technological research, 143
telephone statistics, 206
television viewing public, 79
time series, application to sampling, 50
time study, 141
Tippett, L. H. C., 148
Todd, G. F., 147
training in quality control, 142
Trenaman, J., 212
trouble-shooting, 140

Ungerson, B., 68

van Rest, E. D., 64, 67
variability, method of expression, 207
Vernon, M. D., 209
Vincent, P. D., 70

Wadsworth, R. N., 78, 169
Wainwright, C., 37
Wilkins, L. T., 27, 61, 80
Wright, B. M., 139

Youden, W. J., 148

Printed in Great Britain at THE KYNOCH PRESS, *Birmingham*